NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

LINEAR OPTIMIZATION AND IMAGE RECONSTRUCTION

by

Christopher A. Rhoden

June 1994

Thesis Advisor: Van Emden Henson

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE June 1994	3. REPORT TYPE AND DATES COVERED Master's Thesis

4. TITLE AND SUBTITLE LINEAR OPTIMIZATION AND IMAGE RECONSTRUCTION	5. FUNDING NUMBERS
6. AUTHOR(S) RHODEN, Christopher A.	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey CA 93943-5000	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

12a. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.	12b. DISTRIBUTION CODE A

13. ABSTRACT *(maximum 200 words)*

The Simplex algorithm, developed by George B. Dantzig in 1947 represents a quantum leap in the ability of applied scientists to solve complicated linear optimization problems. Subsequently, its utility in solving finite models, including applications in transportation, production planning, and scheduling, have made the algorithm an indispensable tool to many operations researchers.

This thesis is primarily an exploration of the simplex algorithm, and a discussion of the utility of the algorithm in unconventional optimization problems. The mathematical theory upon which the algorithm is based and a general description of the algorithm are presented. The reader is assumed to have little exposure to *convexity, duality,* or the Simplex algorithm itself. More important to the thesis are the examples that accompany the discussion of the Simplex algorithm. Herein are a variety of unusual applications for the algorithm, including applications in infinite dimensional vector spaces, uniform approximation, and computer assisted tomographic image reconstruction. These examples serve both to facilitate a better understanding of the algorithm, and to present it in unusual settings.

14. SUBJECT TERMS Linear Optimization, Semi-Infinite Linear Programming, Simplex, Convexity, Duality, Image Reconstruction			15. NUMBER OF PAGES 110
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

Approved for public release; distribution is unlimited

UNCLASSIFIED

Linear Optimization and Image Reconstruction

by

Christopher A. Rhoden
Lieutenant, United States Navy
B.A., Miami University, 1987

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN APPLIED MATHEMATICS

from the

NAVAL POSTGRADUATE SCHOOL
June 1994

Author: ______________________
Christopher A. Rhoden

Approved By: ______________________
Van Emden Henson, Thesis Advisor

____
Robert F. Dell, Second Reader

____
Richard Franke, Chairman
Department of Mathematics

ABSTRACT

The Simplex algorithm, developed by George B. Dantzig in 1947 represents a quantum leap in the ability of applied scientists to solve complicated linear optimization problems. Subsequently, its utility in solving finite models, including applications in transportation, production planning, and scheduling, have made the algorithm an indispensable tool to many operations researchers.

This thesis is primarily an exploration of the simplex algorithm, and a discussion of the utility of the algorithm in unconventional optimization problems. The mathematical theory upon which the algorithm is based and a general description of the algorithm are presented. The reader is assumed to have little exposure to *convexity, duality,* or the Simplex algorithm itself. More important to the thesis are the examples that accompany the discussion of the Simplex algorithm. Herein are a variety of unusual applications for the algorithm, including applications in infinite dimensional vector spaces, uniform approximation, and computer assisted tomographic image reconstruction. These examples serve both to facilitate a better understanding of the algorithm, and to present it in unusual settings.

CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

My sincere and humble thanks go to the many who have endured the thesis process in ways more difficult than my own. In particular, thank you to my wife, Marcia, who in spite of our imminent return to sea service, bore the greatest burden in the completion of this thesis. She has unselfishly assumed all of the *important* responsibilities, namely Christopher, Nathanael, and Abigail, over the past months, and sensed perfectly when I needed her to need me at home.

Thanks go to the professionals at the NPS library, who were always helpful, and were particularly patient as I never returned an item on time. Similarly, I express my sincere gratitude for the patience displayed by Gary Krahn and Bruce Robinson, who in spite of the imminence of a dissertation deadline, were always availible for answers.

I remain ever indebted to Professor Maurice Weir. It is he, above all others, who has sparked in me a thirst for mathematical knowledge, and an even greater thirst to understand that knowledge in the context of *absoluteness*.

Finally, my thanks go to Professor Van Henson, who first allowed me to dream big math dreams, then woke me in time for Spring (and though it was never spoken, insisted, thankfully, that softball is a four quarter sport).

I. OVERVIEW OF THESIS

A. THESIS OBJECTIVES

The development of the *S nplex algorithm* by Dantzig in middle of this century represents a milestone in linear optimization techniques. The impact of Dantzig's work is profound. Results of his work include the revival or introduction of a number of mathematical disciplines, including convexity and duality theories. Applications for the Simplex algorithm, and the accompanying refinements, are vast, and many continue to explore new and diverse applications.

The majority of the research on linear optimization problems is taking place in various fields of Operations Research. Of course, the Simplex algorithm itself is particularly well suited to problems in that particular discipline, rendering rapid solutions to production planning models, transportation problems, and a variety of other "real world" applications. A great deal of work was done up to the early 1970's in attempts to mold the Simplex algorithm into an engineering and theoretical mathematical tool. With the advent of more sophisticated computer hardware and software, there may be utility in reconsidering the role of the Simplex algorithm in control, approximation, and other infinite dimensional applications.

This document is intended to serve two main purposes. First, the thesis is intended to serve as an introduction to linear optimization and to the Simplex algorithm, or a theoretical review for readers already familiar with these topics. Second, it is intended to present less traditional problems in a manner that is suitable for solution with the Simplex algorithm.

B. THESIS FORMAT

The thesis is broken into three parts. The first part, consisting of the first two chapters, is devoted to describing sample problems with which the theory of the Simplex algorithm is illustrated. Also image reconstruction is introduced, a problem whose solution by the Simplex algorithm highlights the thesis. These examples are more fully developed in the latter sections.

The first example is particularly unusual, as we find an orthogonal basis of the infinite dimensional vector space $L^2[0,1]$. To the author's knowledge, this is the first attempt to use the Simplex algorithm in this capacity. The formulations that result from this problem are particularly easy to understand, and lend a great deal of understanding to concepts underlying the Simplex algorithm.

The second example may be found infrequently in literature on linear optimization. We seek the best approximation to the exponential function over a closed interval in the uniform norm sense. That is, we formulate a uniform approximation problem as a linear optimization problem. The formulation is used primarily in the discussion of duality.

The final example is again a novel one. We formulate the problem of computer assisted tomographic (CAT) image reconstruction as a linear optimization problem, and solve a small sample problem with the Simplex algorithm.

The second portion, consisting of Chapters IV and V, introduces the machinery behind the Simplex algorithm, culminating with a brief introduction to the algorithm itself. Chapter IV is an exploration of convexity, both as it pertains to sets and functions. The major emphasis of the chapter is on convex subsets of $\mathcal{R}^n$. Chapter V builds on the convexity results as they pertain to duality. Fundamental concepts of duality are presented in this chapter, and it concludes with a generic description of the algorithm.

The thesis concludes with the formulation of the image reconstruction problem as a linear optimization problem in the general case. The first portion of the chapter is devoted to the formulation, followed by the statement of the dual problem. Finally, a sample problem is solved, and some analysis of the appropriateness of the Simplex algorithm as a solution tool for this particular problem is offered.

II. PRELIMINARIES

A. OVERVIEW

We devote this chapter to the preliminaries of linear optimization. We begin by defining three very different examples, which we develop as a means to explore linear optimization methods. We then define the optimization problem in general, and the linear optimization problem specifically. We close with a synopsis of the assumptions that characterize the linear optimization problem.

B. FIRST EXAMPLES

This thesis extensively discusses three examples. We begin by stating two of our three examples to which we refer throughout the thesis. Because of its complexity and importance to this work, the third example is treated separately.

1. Example 1: Generation of an Orthogonal Basis for $L^2[0,1]$

Our first example is one of importance in many areas of approximation. We wish to find some orthogonal basis for an infinite dimensional vector space. The utility of such bases may be found in any elementary linear algebra or applied mathematics text. The interested reader is referred to [Ref. 1]. The specific vector space with which we are concerned is the space of functions defined by

$$L^2[0,1] = \left\{ f : \|f\| = \left(\int_0^1 f(x)^2 dx \right)^{\frac{1}{2}} < \infty \right\}.$$

We note that the above norm is induced by the inner product,

$$\langle f, g \rangle = \int_0^1 f(x)g(x)\, dx.$$

That is,

$$\|f\| = \sqrt{\langle f, f \rangle}.$$

In particular, we seek an orthogonal polynomial basis, and derive an optimization technique to find a polynomial p_n, of order n, when we are given a set of orthogonal polynomials, $p_0, p_1, \ldots, p_{n-1}$. Recursive application of a method for generating p_n leads to a complete set of basis polynomials. The polynomial basis is of particular importance, as the *Weierstrass Approximation Theorem* assures us that any continuous function f, defined on $[0,1]$, may be approximated arbitrarily well with polynomials [Ref. 2].

There are a number of existing techniques for the generation of orthogonal polynomials. For example, the Gram-Schmidt algorithm may be applied to the sequence $\{1, x, x^2, \ldots, x^n, \ldots\}$. Another approach involves solving a three-term recurrence that generates the polynomials. We consider an optimization approach, in which we formulate an optimization problem whose solution gives us p_n. It is an approach that is suitable for inductive iteration.

2. Example 2: Uniform Approximation of the Exponential Function

The second example is a specific problem in uniform approximation. We seek the linear combination of polynomials on the interval $[0,3]$ that best approximates the exponential function in the uniform norm sense. The problem, consequently, is to find the coefficients α_i, that minimize the expression

$$\max_{t\epsilon[0,3]} \mid f(t) - e^t \mid,$$

$$\text{where } f(t) = \sum_{i=0}^{n} \alpha_i t^i.$$

We consider specific cases of this example. That is, we seek the polynomial for some fixed degree, n, that best approximates the exponential function. Note that the uniform approximation problem is fundamentally an optimization problem. The use of the Simplex algorithm to solve the problem is, however, unusual.

C. EXAMPLE 3: THE IMAGE RECONSTRUCTION PROBLEM

The third example is the image reconstruction problem. As with the first two examples, there are many existing techniques for solving this problem. Unlike the others, however, this is an active area of modern research, and the best methods of solution may yet be unknown. The reader is referred to [Ref. 3] for a thorough treatment of the problem, and to [Ref. 4] and [Ref. 5] for an introduction to some recently developed solution techniques.

Suppose a neurosurgeon wishes to rule out the possibility that a patient, Fred, suffers from a brain tumor. Further, the physician opts to make use of the CAT (Computer Aided Tomography) scan device, and examine the inside of Fred's head without exploratory surgery.

The CAT scan machine works by projecting a finite number of X-rays of known intensity into tae patient's head from a finite number of positions. The intensity of the X-rays upon leaving Fred's head is measured. The intensity of the emergent X-ray depends essentially on the density of Fred's head over the locations through which the X-ray passes. Having collected data from a number of X-rays, the gathered data are processed, forming a model of the density of Fred's head. That is, the processing of the data results in the construction of an image, and presumably, an image that closely corresponds to the interior of the Fred's head. This data processing, in this example, constitutes solving the image reconstruction problem.

1. X-Ray Computed Tomography

Understanding the methods of reconstruction requires that we know the process by which the data for reconstruction are obtained. We begin with a basic discussion of the manner in which an X-ray moves through an object of homogeneous density, then derive the manner in which it moves through more complicated media.

It has been shown empirically that the fractional decrease in beam intensity of a narrow beam of X-ray photons passing through a homogeneous material

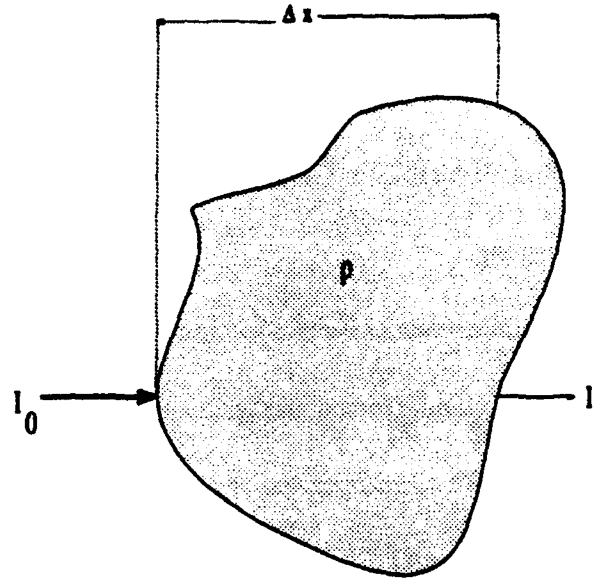

Figure 1. *An X-Ray Over a Homogeneous Object:* I_0 = Input Intensity. I = Emergent Intensity. ρ = Density.

is given by the relationship ([Ref. 6])

$$\frac{I}{I_0} = e^{-\rho(\Delta x)},$$

where I_0 is the X-ray input intensity, and I is the observed intensity after the ray passes a distance Δx through the material. See Figure 1. The parameter ρ is determined by the density of the material.[1] For two media, the fractional decrease is, predictably,

$$\frac{I}{I_0} = e^{-(\rho_1(\Delta x_1)+\rho_2(\Delta x_2))},$$

where Δx_i denotes the distance the X-ray travels through the i^{th} medium.

---

[1] ρ also depends, to a lesser extent, on a number of other factors, including the nuclear composition characterized by the atomic number Z, a function of th· ·mogeneous material. [Ref. 3] pertains. For the purposes of this paper, the effects of other paran· ·ters are assumed to be nil.

Let us partition the media through which the narrow beam travels into n homogeneous segments. Denote the density over a single segment by $\rho(x)$. The decrease in this case is expressed by

$$\frac{I}{I_0} = \exp\left(-\sum_{i=1}^{n}\rho(x_i)\Delta x_i\right). \tag{II.1}$$

Letting $n \Rightarrow \infty$, and $\Delta x_i \Rightarrow 0$, equation (II.1) becomes

$$\begin{aligned}\frac{I}{I_0} &= \lim_{n\to\infty,\Delta x\to 0}\exp\left(-\sum_{i=1}^{n}\rho(x_i)\Delta x_i\right)\\ &= \exp\left(-\int\rho(x)dx\right),\end{aligned}$$

implying

$$-\ln\frac{I}{I_0} = \int\rho(x)dx. \tag{II.2}$$

Concluding, let l be the line describing the path of the X-ray, and the function, $f(x,y)$ is the density of the media over l. Let ds denote a length over the line l. Equation (II.2) may be written in the form

$$-\ln\frac{I}{I_0} = \int_l f(x,y)ds. \tag{II.3}$$

2. The Radon Transform

This section is an introduction to the Radon Transform, and elaborates its relation to the data collected with the X-ray. We first define the transform, then briefly describe some of its properties. The discussion in this section pertains to the two-dimensional case. That is, we wish to find the density of an object defined over some subset of $\mathcal{R}^2$. For generalizations into higher dimensions, see [Ref. 3].

We begin by considering some density function f, defined and bounded on a simply connected, compact subset $\Omega \subset \mathcal{R}^2$. Define L to be the set of all straight lines passing through any portion of Ω. That is, $L = \{l : l \cap \Omega \neq \emptyset\}$. Note that the cardinality of L is uncountably infinite. The *Radon transform* is defined by all possible line integrals of the form:

$$\tilde{f} = \int_{l_j} f(x,y)\, ds, \quad j \in J, \tag{II.4}$$

where ds is an increment of length along l_j, and J is the index set of the set L.

Consider how the lines, over which the integrals above are computed, are determined. Let $\mu = [\cos\phi, \sin\phi]^{\mathrm{T}}$. Then for a fixed angle of rotation ϕ and a distance ρ from the origin, we may identify each line, l_i by the set of vectors, $\mathbf{x} = [\mathrm{x}, \mathrm{y}]^{\mathrm{T}}$, that satisfy the equation

$$\langle \mathbf{x}, \mu \rangle = \mathrm{x}\cos\phi + \mathrm{y}\sin\phi = \rho.$$

(See Figure 2). Consequently, we may denote each of the line integrals defining the Radon transform by

$$\tilde{f}(\phi, \rho) = \int_{\langle \mathbf{x}, \mu \rangle = \rho} f(\mathbf{x})\, d\mathbf{x}. \tag{II.5}$$

Again, it is vital to note that the Radon transform is defined by the collection of *all* such line integrals. Consequently, to determine the Radon transform fully, we must know $\tilde{f}(\phi, p)$ for all values of ϕ and p. When we know the value of the line integrals for only certain values of ϕ and p, we say that we have a *sample* of the transform.

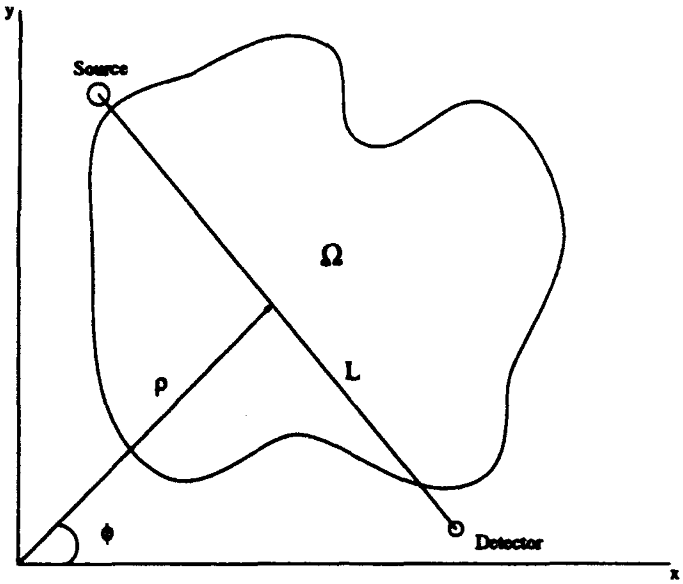

Figure 2. *The Line,* L, *as it Relates to* (ρ, ϕ)

3. The Problem Statement

We note that the right hand sides of Equations (II.3) and (II.4) are identical. We conclude, then, that if the X-ray is sufficiently narrow, and we are able to take an X-ray along all possible lines, the resultant infinite collection of data corresponds to the Radon transform of the desired density function.

The Radon transform has been shown to be one-to-one ([Ref. 3]). That is, when *all* values of the line integrals are known, one may determine the unique density that produces the observed transform data. However, in most cases of practical interest, we are presented with but a sample of the transform from which to reconstruct an image. That is, we are able to collect only a finite number of X-rays. Additionally, the photon beam is not sufficiently narrow to be a true line integral defining the transform. In this case, inverting the transform is an ill-posed problem.

If there exists one density function whose sampled Radon transform equals a given data set, then there exist infinitely many density functions, f such that $\tilde{f} = \mathbf{b}$, where $\mathbf{b}$ is the data obtained from a transform sample. It is this fact that leads us to investigate an optimization approach to the image reconstruction problem.

D. OPTIMIZATION

Each of the examples can be formulated as an optimization problem. Fundamental to any optimization problem, and to the *Linear Optimization Problem*, in particular, are the concepts of *feasible set* and *objective function*.

1. Feasible Sets

To help explain a feasible set, we consider an example. Suppose we wish to model the production schedule for a baseball and softball manufacturing plant. The company is required to make at least 500 baseballs and 1000 softballs each month to satisfy contractual agreements. The company expects to procure 2,000 pounds of stuffing material, and 3,000 square feet of leather covers. Each baseball requires $\frac{1}{4}$ pound of stuffing, and $\frac{1}{2}$ square feet of leather. The requirements for the softballs are $\frac{3}{8}$ pounds and $\frac{3}{4}$ square feet of stuffing and leather respectively. Then of all possible production schedules, we restrict our attention to those that fulfill contractual requirements and do not utilize assets which are not available. Let b and s be the number of baseballs and softballs, respectively, produced in a month. Then we require that:

$$b \geq 500$$

$$
\begin{aligned}
s &\geq 1{,}000 \\
\frac{1}{4}b + \frac{3}{8}s &\leq 2{,}000 \\
\frac{1}{2}b + \frac{3}{4}s &\leq 2{,}000. \qquad \text{(II.6)}
\end{aligned}
$$

We have defined a subset of all possible schedules by a group of mathematical relationships. In this example, the feasible set is the set of all production schedules that satisfy the equations of (II.6). In general, we define the feasible set, Y, to be the collection of values satisfying the mathematical relationships imposed by the problem.

2. The Objective Function

The objective function, g, defined over a feasible set, Y, is the function by which one models the quality of a solution. In the manufacturing schedule example, we might logically define the objective function to be profit. Supposing that each baseball contributes \$1 of profit, and each softball, \$.75, we could write our objective function:

$$g = b + .75s,$$

and we seek the maximum value of g over Y.

Simply stated, an optimization problem is expressed by "Considering all members of the feasible set, Y, which member(s) results in the optimum value of the objective function, g?"

E. LINEAR OPTIMIZATION

The *Linear Optimization Problem,* or *LOP,* is defined by the criteria that the objective function and the relatic nships defining the feasible set be linear in our *decision variables,* or the variables representing the values we seek. Then we may write the *LOP* as follows:

> Let a vector $\mathbf{c} = [c_1, c_2, \ldots, c_n]^T \in \mathcal{R}^n$, a non-empty index set S, and for every $s \in S$ a vector $\mathbf{a}(s) \in \mathcal{R}^n$, and a real number $b(s)$ be given. Defining $\langle \mathbf{u}, \mathbf{v} \rangle$ as the standard inner product, we seek a vector $\mathbf{y} \in \mathcal{R}^n$, called the *optimal vector*, that minimizes:
>
> $$\langle \mathbf{c}, \mathbf{y} \rangle$$
>
> while satisfying:
>
> $$\langle \mathbf{a}(s), \mathbf{y} \rangle \geq b(s),$$
>
> for all $s \in S$.

We observe that a linear maximization problem may be put into the form above in the following way. The linearity of the objective function assures us that it is continuous on Y, and that the feasible set is compact. Then $\max(f) = \min(-f)$, and we may equivalently seek to minimize $(-f)$.

A similar change may be made in the constraints to reverse inequalities if necessary. That is, the problems

$$\begin{array}{lll} \text{Maximize:} & \langle \mathbf{c}, \mathbf{y} \rangle & \\ \text{Subject to:} & \langle \mathbf{a}(s), \mathbf{y} \rangle & \leq b(s) \\ & \text{for all } s \in S & \end{array}$$

and

$$\begin{aligned} \text{Minimize:} \quad & -\langle \mathbf{c}, \mathbf{y} \rangle \\ \text{Subject to:} \quad & -\langle \mathbf{a}(\mathrm{s}), \mathbf{y} \rangle \geq -b(s) \\ & \text{for all } s \in S \end{aligned}$$

are identical.

1. The Linear Program

The case where the cardinality of $S = m < \infty$ defines a *Linear Program.* This special case of the LOP is of particular interest as it forms the basis for finding solutions to LOPs when the index set S is infinite. Throughout this thesis, the reader may assume that discussion of the general linear optimization problem permits the possibility of an infinite index set S, unless explicity otherwise noted.

Now, however, we examine this Linear Programming case to clarify the concept of linearity. The problem becomes

$$\begin{aligned} \text{minimize} \quad & \langle \mathbf{c}, \mathbf{y} \rangle \\ \text{subject to:} \quad & \langle \mathbf{a}(\mathrm{s_i}), \mathbf{y} \rangle \geq b(s_i) \\ \text{for } i = 1, 2, \ldots, m, \quad & \text{over all} \quad \mathbf{y} \in \mathcal{R}^{\mathrm{n}}. \end{aligned} \tag{II.7}$$

Let $a_j(s_i)$ denote the j^{th} component of the vector $\mathbf{a}(\mathrm{s_i})$. We may write the problem as

$$\text{Minimize} \qquad c_1 y_1 + c_2 y_2 + \cdots + c_n y_n$$

$$
\begin{aligned}
\text{Subject to:} \quad & a_1(s_1)\, y_1 + a_2(s_1)\, y_2 + \cdots + a_n(s_1)\, y_n \geq b(s_1) \\
& a_1(s_2)\, y_1 + a_2(s_2)\, y_2 + \cdots + a_n(s_2)\, y_n \geq b(s_2) \\
& \vdots \\
& a_1(s_m)\, y_1 + a_2(s_m)\, y_2 + \cdots + a_n(s_m)\, y_n \geq b(s_m)
\end{aligned}
$$

over all $y \in \mathcal{R}^n$. (II.8)

We note that in this case, we may define the feasible set by the notation

$$A^T \mathbf{y} \geq \mathbf{b} \tag{II.9}$$

with $A \in \mathcal{R}^{n \times m}$, and the i^{th} column of A is $\mathbf{a}(s_i)$. The i^{th} component of the vector $\mathbf{b}$ is given by $b(s_i)$. The linearity assumptions can be expressed, as follows [Ref. 7].[2]

1. *Proportional* : The objective function is linear in the feasible set, Y, in the following sense. Given a variable, y_j, its contribution to the objective function is $c_j y_j$. So then a change of d units in y_j results in a change in the objective function value of $c_j d$. Similarly, the constraints are linear with respect to the variable y_j, insofar as the contribution of the variable y_j to the i^{th} constraint is $a_j(s_i) y_j$. Then changing the value of y_j by d units changes the value of the left-hand-side of the i^{th} constraint by $a_j(s_i) d$ units.

2. *Deterministic* : The components of the vectors $\mathbf{a}(s)$ and $\mathbf{c}$ are all determined, as is each scalar $b(s)$. That is, if the components are derived from some

---

[2] [Ref. 7] also identifies the qualities of *additivitiy* and *divisibility* as requirement of the linear optimization problem. These qualities are deemed to be inherent in the qualities defined above.

stochastic model, their variability is disregarded, and the numbers are fixed for a given linear optimization problem. Having defined the Linear Optimization problem, we now turn our attention to exploring the utility of solution techniques to non-traditional optimization problems.

III. THE EXAMPLES, A DIFFERENT PERSPECTIVE

A. OVERVIEW

This section addresses some of the basic properties of the sets from which we choose an optimal vector in our examples. It is the structure which we are able to assign to these sets that permits us to exploit the theories regarding convexity, and subsequently, the duality results which we derive in subsequent chapters. We then introduce assumptions that refine the feasible sets.

B. LINEAR VECTOR SPACES

Before proceding to the specific examples, we first turn our attention to the matter of *linear vector spaces.* A vector space, L, is called a linear vector space if for any vectors $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathrm{L}$ and any real scalars α and β the following results hold [Ref. 1]:

1. $\alpha(\mathbf{x}+\mathbf{y}) = (\alpha\mathbf{x} + \alpha\mathbf{y}) \in \mathrm{L}$
2. $\alpha(\beta\mathbf{x}) = (\alpha\beta)\mathbf{x}$
3. $\mathbf{x}+\mathbf{y} = \mathbf{y}+\mathbf{x}$
4. $\mathbf{x}+(\mathbf{y}+\mathbf{z}) = (\mathbf{x}+\mathbf{y})+\mathbf{z}$

For each of the example problems, the feasible set is a subset of a linear vector space. Consider the problem of finding an orthogonal polynomial, p_n, of order n. It is elementary that the set of polynomials of order n form a linear vector space. The same holds for the problem of finding the polynomial that best approximates the exponential function on $[0,3]$. Finally, in example three, we have specified that we wish to find a density function, f, from the set of all bounded, piecewise continuous functions with support over a compact set Ω. The set of all such functions is a linear vector space.

Equally important to our discussion is the concept of a norm. In general, a norm on a linear vector space L is defined to be a mapping, denoted $\|\dot{}\| : L \rightarrow \mathcal{R}^+$ satisfying the following rules [Ref. 1]. For all $\mathbf{x}, \mathbf{y} \in \mathrm{L}$, and $\alpha \in \mathcal{R}$,

1. $\|\mathbf{x}\| \geq 0$ and $\|\mathbf{x}\| = 0, \Leftrightarrow \mathbf{x} = 0$
2. $\|\alpha x\| = \mid \alpha \mid \|\mathbf{x}\|$,
3. $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$.

Any linear vector space equipped with such a function is said to be a *normed linear vector space.* Each of the feasible sets of the examples is a subset of a normed linear vector space. The first two examples are clearly so. Any norm on $\mathcal{R}^n$ suffice. In the third example, we use the infinity norm, defined by:

$$\|f\|_\infty = \sup_{\omega \in \Omega} \mid f(\omega) \mid$$

as an appropriate norm.

C. REFINING THE FEASIBLE SUBSET OF THE ORTHOGONAL BASIS PROBLEM

In the first example, we are interested in finding a polynomial, p_n, of order n, such that:

$$\langle p_n, p_i \rangle = \int_0^1 p_n p_i dx = 0,$$
$$\text{for all } 0 \leq i \leq n-1,$$

where the result is assumed true for all $p_i, p_j, i \neq j$. That is, given orthogonal polynomials $p_0, p_1, \ldots, p_{n-1}$, we seek a polynomial of order n, orthogonal to all of the polynomials of lower order. We formulate this problem:

$$\begin{array}{lll} \text{minimize:} & \sum_{i=0}^{n-1} \int_0^1 p_n p_i dx & \\ \text{Subject to:} & \int_0^1 p_n p_i \geq 0, & \text{for } i = 1, 2, \ldots, n-1. \end{array} \qquad \text{(III.1)}$$

Theorem 1. *The optimal objective function value for the orthogonal polynomial problem is zero, and any optimal vector, p_n satisfies the desired orthogonality conditions.*

Proof: Since we know triangular families of orthogonal polynomials exist, we conclude immediately that the optimal objective function value is bounded above by zero. The constraint gives us zero as a lower bound. That any optimal vector satisfies our orthogonality conditions is immediate from these facts. That is, a zero objective function value, in conjunction with the constraint ensures orthogonality. □

There are infinitely many polynomials that satisfy the above criteria. Specifically, if the objective function evaluates to zero for some p_n, it clear evaluates to zero for $\alpha\, p_n$, for any $\alpha \in \mathcal{R}$. Consequently, we add the additional constraint that the polynomial we desire is the monic orthogonal polynomial. The additional constraint leads rather easily to an $n \times n$ linearly independent system of inequalities, where the unknown element of $\mathcal{R}^n$ is the vector whose components are the coefficients of the desired polynomial.

To illustrate, let us consider the specific cases of finding the first order and second order polynomial satisfying (III.1). We input the zeroth order monic polynomial, $p_0 = 1$, to start the iterative process.

In the first order case, the objective function

$$\sum_{i=0}^{n-1} \int_0^1 p_n(x)\; p_i(x) dx$$

is simply

$$\int_0^1 p_1(x) dx = \int_0^1 (x+\alpha)\; dx = \frac{1}{2} + \alpha.$$

The optimization problem takes on the form,

$$\begin{aligned} \text{minimize:} \quad & \tfrac{1}{2} + \alpha \\ \text{subject to:} \quad & \tfrac{1}{2} + \alpha \;\; \geq 0, \end{aligned}$$

from which we observe that $\alpha = -\frac{1}{2}$, and conclude that $p_1(x) = x - \frac{1}{2}$. While the solution of this particular problem is trivial, there are some important conceptual principles working here. Considering the problem in terms of the linear optimization problem, observe that the feasible set is the set of all real numbers α, with $\alpha \geq -\frac{1}{2}$.

As the function we seek to minimize takes on the form, $C + \alpha$, where C is a fixed c .istant, we clearly wish to select the smallest possible value for α.

Similarly, consider the formulation of the problem of finding the second order polynomial. We define the polynomials p_0 and p_1, as above, and let $p_2 = x^2 + \alpha_1 x + \alpha_0$. Computing the integrals, we find that

$$\int_0^1 p_0 p_2 = \int_0^1 p_2 = \frac{1}{3} + \frac{\alpha_1}{2} + \alpha_0,$$

and

$$\begin{aligned}
\int_0^1 p_1 p_2 &= \int_0^1 \left(x - \frac{1}{2}\right)\left(x^2 + \alpha_1 x + \alpha_0\right) dx \\
&= \int_0^1 x^3 + \left(\alpha_1 - \frac{1}{2}\right) x^2 + \left(\alpha_0 - \frac{\alpha_1}{2}\right) x - \frac{\alpha_0}{2} \\
&= \frac{1}{4} + \left(\frac{\alpha_1}{3} - \frac{1}{6}\right) + \left(\frac{\alpha_0}{2} - \frac{\alpha_1}{4}\right) - \frac{\alpha_0}{2} \\
&= \frac{1}{12} + \frac{\alpha_1}{12}.
\end{aligned}$$

Then the linear optimization problem

$$\begin{aligned}
\text{minimize:} &\quad \textstyle\sum_{i=0}^{n-1} \int_0^1 p_i p_n \\
\text{subject to:} &\quad \textstyle\int_0^1 p_i p_n \geq 0, \quad i = 1, 2, \ldots, n-1,
\end{aligned}$$

becomes:

$$\begin{aligned}
\text{minimize:} &\quad \tfrac{5}{12} + \tfrac{7}{12}\alpha_1 + \alpha_0 \\
\text{subject to:} &\quad \tfrac{1}{3} + \tfrac{1}{2}\alpha_1 + \alpha_0 \quad \geq 0, \\
&\quad \tfrac{1}{12} + \tfrac{1}{12}\alpha_1 \qquad \geq 0.
\end{aligned} \tag{III.2}$$

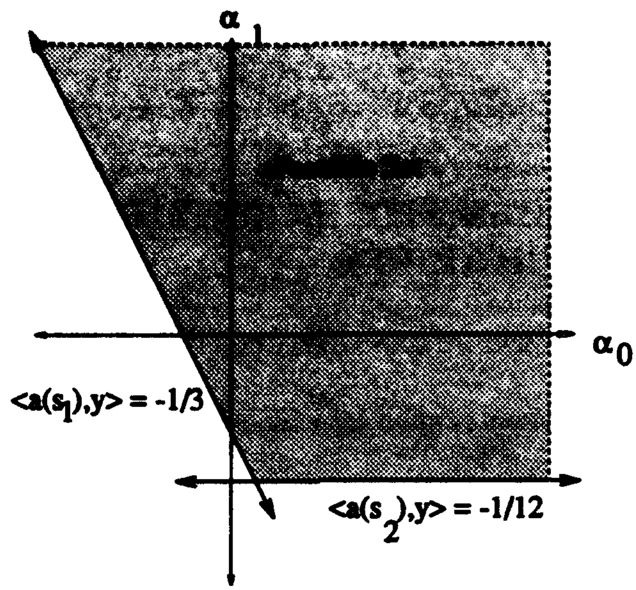

Figure 3. *The Feasible Set of* Example 1*: n=2*

As we are currently finding the feasible set, and viewing the problem in terms of the general formulation, we make the following observations. The index set S has cardinality 2. By rearranging the constraint equations, we find the constraint vectors are

$$\begin{aligned} \mathbf{a}(s_1) &= \begin{bmatrix} \frac{1}{2} & 1 \end{bmatrix}^T \\ \mathbf{a}(s_2) &= \begin{bmatrix} \frac{1}{12} & 0 \end{bmatrix}^T, \end{aligned}$$

with $b(s_1) = -\frac{1}{3}$, and $b(s_2) = -\frac{1}{12}$. As the vector we seek, $\mathbf{y} = [\alpha_1 \quad \alpha_0]^{\mathrm{T}} \epsilon\, \mathcal{R}^2$, we may illustrate the feasible set as in Figure 3.

We observe that we have a problem of finding the optimal vector in $\mathcal{R}^2$ when seeking the second order orthogonal polynomial. This property generalizes for any order of polynomial. That is, if we seek a polynomial of order n, we seek a vector, $\mathbf{y} \in \mathcal{R}^n$, giving the coefficients for the optimal monic polynomial.

D. THE FEASIBLE SET IN THE UNIFORM APPROXIMATION PROBLEM

Consider the problem of approximating the exponential function, e^t, in the interval $[0, 3]$ by a linear combination of polynomials. We have specified that we wish to find the combination that minimizes the maximum residual over the interval, and not the total residual. Hence, we are not solving the least squares problem, where orthogonality of the approximating functions dramatically simplifies the task. With the uniform approximation problem, however, orthogonality of the polynomials is not particularly useful. Therefore, rather than using the orthogonal polynomials above, we merely specify the degree of the approximating polynomial. Thus we seek a linear combination of the polynomials

$$\sum_{i=0}^{n} \alpha_i P_i(t),$$

$$\text{where} \qquad P_i(t) = t^i, \qquad i = 0, 1, 2, \ldots n.$$

Consider the specific example for $n = 1$. We seek a polynomial

$$\langle \mathbf{T}, \mathbf{y} \rangle, \text{ where } \mathbf{T} = [1, \mathrm{t}]^{\mathrm{T}}, \text{ and } \mathbf{y} = [\alpha_0, \alpha_1]^{\mathrm{T}} \epsilon \mathcal{R}^2.$$

Since the vector $\mathbf{T}$ is fixed, the problem is equivalently one of determining the optimal vector, $\mathbf{y} \in \mathcal{R}^2$. We summarize with a preliminary statement of the problem.

$$\text{minimize:} \qquad \max_{t \epsilon [0,3]} \left| \left(\sum_{i=0}^{n} \alpha_i t^i \right) - e^t \right|$$

$$\text{over all } \mathbf{y} = [\alpha_0, \ldots, \alpha_n]^T \, \epsilon \, \mathcal{R}^{n+1}.$$

Observe that the objective function is non-linear in the decision variables, α_i, $i = 1, \ldots, n$. Also observe that the feasible set is $\mathcal{R}^{n+1}$ in its entirety. That is, any combination of real coefficients is feasible, since there are currently no constraints.

E. CONVENTIONS OF IMAGE RECONSTRUCTION

For Example 3, we have specified that we wish to find some function, f, defined on the simply connected, compact set $\Omega \subset \mathcal{R}^2$. Assume that Ω is a circle of radius 1. We also assume that the function that we seek is piecewise continuous on Ω. The piecewise continuity restriction is justified by the physical nature of the problem we are solving. We call the space of such functions F. Here it is useful to define a basis for F, and we select a logical basis in view of the problem we wish to solve.

As we have stated, the the formal inverse of the Radon transform is well defined. Our difficulty results from our inability to compute the uncountably infinite number of line integrals defining the Radon transform. This difficulty stems first from the fact that the region over which an X-ray is measured is not one-dimensional. That is, the region over which the X-ray measures mass has both width and length. Each X-ray measures the density of the medium over some strip, as in Figure 4. Additionally, the number of data points from which one reconstructs an image is

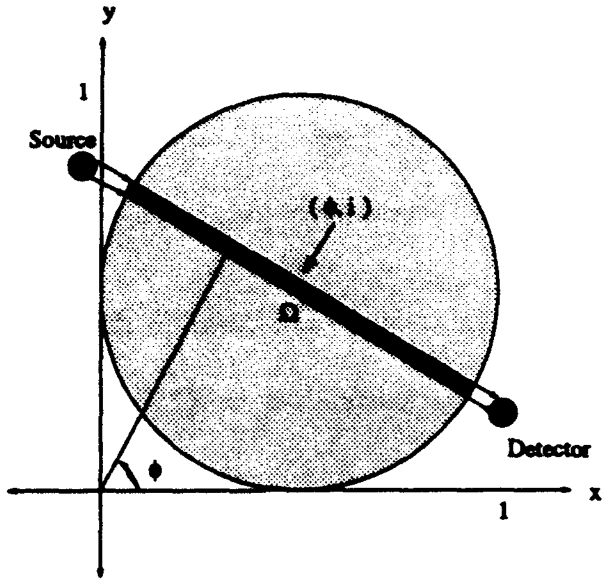

Figure 4. *A Single Density Measuring Strip*

finite, rather than uncountably infinite, as required for formal transform inversion. A more accurate perspective from which to view the data obtained by the X-rays is presented here.

Begin by fixing an angle ϕ. We associate with each strip of ϕ, a label (ϕ, i). We introduce the strip characteristic function, γ. Define the real valued function γ defined on Ω by the rule

$$\gamma_{\phi,i}(\omega) = \begin{cases} 1, & \text{if } \omega \text{ lies in strip } (\phi, i) \\ 0, & \text{otherwise.} \end{cases}$$

Then an integral defining the sampled Radon transform, for a fixed angle, ϕ, and a fixed strip, (ϕ, i), becomes

$$\tilde{f}_{\phi,i} = \int_\Omega f(\omega)\, \gamma_{\phi,i}(\omega)\, d\omega.$$

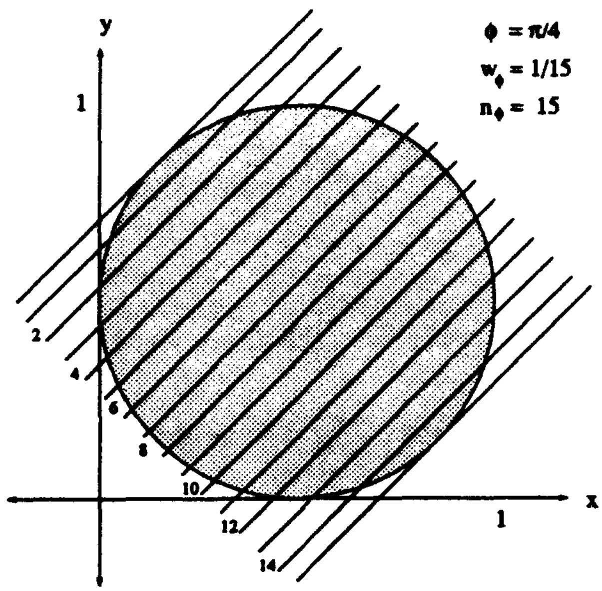

Figure 5. *A Single View*: Note that strips do not overlap, and cover Ω completely.

Let us define a *view* to be the set of all strips for some fixed angle, ϕ. We impose two restrictions. First, we require that all strips of a view are non-overlapping. Mathematically, if (ϕ, i) and (ϕ, j) correspond to two strips of the same view,

$$\gamma_{\phi,i}(\omega)\ \gamma_{\phi,j}(\omega) = 0, \text{ for all } i \neq j,$$

for any $\omega \in \Omega$.

Second, we require that the strips composing a view completely cover the compact set. That is, for any $\omega \in \Omega$. and every angle ϕ, there exist some strip (ϕ, i) such that

$$\gamma_{\phi,i}(\omega) = 1.$$

See Figure 5 for a graphical presentation of these properties.

Assume that we have some manner in which to control the width of the strips. Then we may select some number of strips of equal width for each view. Identifying the number of strips for view ϕ as n_ϕ, and the width of a strip for view ϕ as b_ϕ, we may conclude that:

$$n_\phi \times b_\phi = 1, \text{ the diameter of } \Omega.$$

For a finite number of views, N_v, the application of this convention partition the set Ω into a finite number of polygons. We call the set of these polygons a *polygonal partition* of Ω. Figure 6 illustrates the manner in which these polygons are formed. With each of the resultant polygons, s_j, we associate a scalar, $area(s_j)$, and a characteristic function,

$$\psi_j(\omega) = \begin{cases} 1 & \text{if } \omega \in s_j, \text{ and,} \\ 0 & \text{otherwise.} \end{cases}$$

It is the set of these characteristic functions, ψ_j that we use as the basis for the function space, F.

Theorem 2. *For any continuous function, g defined on Ω, and any $\varepsilon > 0$, there exist some polygonal partition on n polygons, and some function*

$$f = \sum_{i=1}^{n} \alpha_i \psi_i$$

such that $\|f - g\|_\infty < \varepsilon$.

Note that we may write $\|f(\omega) - g(\omega)\|_\infty$ with the equivalent notation,

$$\|f(\omega) - g(\omega)\|_\infty = \max_{j=1,\ldots,n} \left\{ \max_{\omega \in s_j} |\ (f(\omega) - g(\omega))\psi_j\ | \right\}.$$

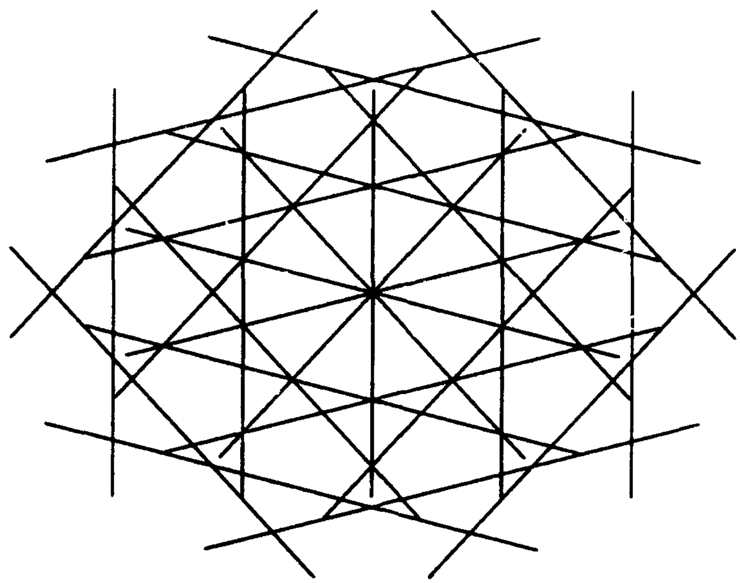

Figure 6. *Polygons Created by the Views* $\frac{i\pi}{5}$, $\{i = 0, 1, \ldots, 4\}$, *Each with 4 Strips* .

One may easily verify that $\|f(\omega) - g(\omega)\|_\infty$ is a norm. Note that we may use the maximum over j rather than the supremum, as the polygonal partition is a finite set. The properties of our function space, F, allow us to use the maximum rather than the supremum over each polygon, s_j.

Proof: Let g be any continuous function in F, and let $\varepsilon > 0$ be given. As g is continuous, there exists some $\delta > 0$ such that

$$\begin{aligned} \|(x, y) - (p, q)\|_\infty &< \delta \text{ implies that} \\ \|g(x, y) - g(p, q)\|_\infty &< \varepsilon. \end{aligned} \tag{III.3}$$

We use only two angles, $\phi_1 = 0$ and $\phi_2 = \frac{\pi}{2}$. Let $n_{\phi_1} = n_{\phi_2} = \lceil \frac{1}{\delta} \rceil$. Note that this implies that

$$n = n_{\phi_1} \times n_{\phi_2} = \left\lceil \frac{1}{\delta} \right\rceil^2,$$

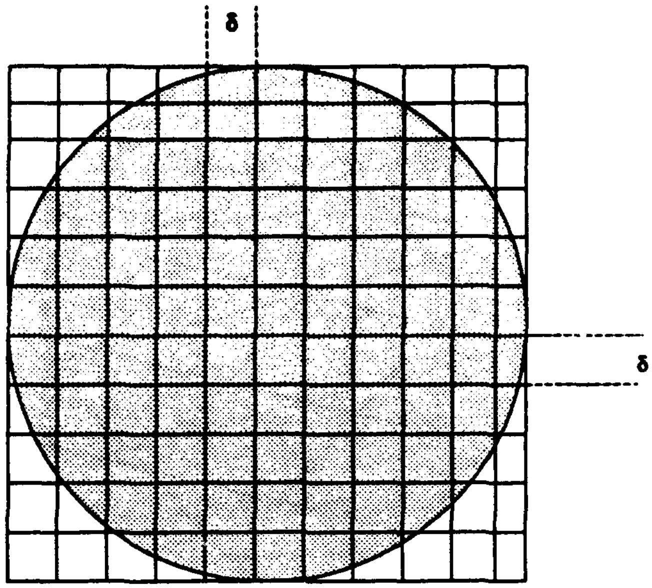

Figure 7. *An Arbitrary Square of the Proof Partition*

as in Figure 7. Further, for any two points, (x, y) and (p, q) in a fixed polygon, $\|(x, y) - (p, q)\|_\infty < \delta$.

Let $f = \sum_{i=1}^{n} \alpha_i \psi_i$. We now consider $\|f - g\|_\infty$.

$$\begin{aligned}
\|f - g\|_\infty &= \max_{j=1,\ldots,n} \left\{ \max_{\omega \epsilon s_j} \mid (f(\omega) - g(\omega))\psi_j \mid \right\} \\
&= \max_{j=1,\ldots,n} \left\{ \max_{\omega \epsilon s_j} \mid \sum_{i=1}^{n} \alpha_j \psi_i \psi_j - g(\omega)\psi_j \mid \right\} \\
&= \max_{j=1,\ldots,n} \left\{ \max_{\omega \epsilon s_j} \mid \alpha_j \psi_j - g(\omega)\psi_j \mid \right\} \\
&= \max_{j=1,2,\ldots,n} \left\{ \max_{\omega \epsilon s_j} \mid \alpha_j - g(\omega) \mid \right\}.
\end{aligned}$$

Since g is continuous and each of our polygons is compact, g achieves its maximum and minimum on each square. For each square, s_j, define $M_j = \max_{\omega \epsilon s_j} g(\omega)$, and $m_j = \min_{\omega \epsilon s_j} g(\omega)$. Choose

$$\alpha_j = \frac{M_j + m_j}{2}.$$

Using the continuity of g to invoke the intermediate value theorem, there exists some $\hat{\omega} \epsilon s_j$ such that $g(\hat{\omega}) = \alpha_j$. Further, we know that $\omega \epsilon s_j \Rightarrow \|\omega - \hat{\omega}\|_\infty \leq \delta$. Therefore, for any square, s_j,

$$\begin{aligned} \max_{\omega \epsilon s_j} \mid g(\hat{\omega}) - g(\omega) \mid &< \varepsilon, \text{ implying} \\ \mid \alpha_j - g(\omega) \mid &< \varepsilon, \text{ for } j = 1, \ldots, n. \end{aligned}$$

Therefore

$$\max_{j=1,\ldots,m} \left\{ \max_{\omega \epsilon s_j} \mid (f(\omega) - g(\omega))\psi_j \mid \right\} < \varepsilon.$$

□

While the above proof uses only two views, one may increase the number of views, or insist on narrower strips in the partition of Ω. Clearly, such a refinement can not degrade the approximation of the function g, but only maintain or improve it. We may, at worst, maintain the same constant values over the new polygons that they were assigned over the coarser partition.

We now demonstrate the utility of defining a basis for F. Let $k = \sum_\phi n_\phi$. That is, let k denote the total number of strips defining sample transform. For any polygonal partition P on n polygons, the sample Radon transform with respect to P, which we denote $\tilde{f}_P$, may be written as

$$\tilde{f}_P = A^T \mathbf{y}$$

where A is an $n \times k$ matrix, and $\mathbf{y} \in \mathcal{R}^n$. The matrix A is given by:

$$A_{ij} = \begin{cases} 0, & \text{if } \gamma_j(\omega)\ \psi_i(\omega) = 0 \text{ for all } \omega \\ area(\mathrm{s_i}), & \text{otherwise.} \end{cases}$$

That is, the A_{ij} represents the area of the i^{th} polygon if the polygon falls within strip j. The i^{th} component of $\mathbf{y}$ is the mean density of the function f over polygon i.

For any fixed polygonal partition, the feasible set is a subset of the infinite dimensional vector space, F. Each element of the subset may be thought of as a vector in $\mathcal{R}^n$. Without further restriction, the feasible set becomes the set of all vectors, $\mathbf{y} \in \mathcal{R}^n$ such that $A^T\mathbf{y} = \tilde{\mathbf{f}}_{\mathrm{P}}$. We exploit many of the subsequent theorems as a result of the ability to translate the problem into $\mathcal{R}^n$.

IV. CONVEXITY

A. OVERVIEW

In this chapter, we investigate the concept of convexity, both as it pertains to sets and to functions. The primary motivation for this investigation comes from the fact that we may, when certain convexity conditions are met, conclude that local maxima and minima are global. Stated differently, we may eliminate a portion, often a large portion, of our feasible set from consideration when attempting to find the optimal value of our objective function. This chapter lays the groundwork for our investigation into duality, contained in the following chapter.

This and the following chapter form the foundation for linear optimization, and, consequently, the concepts and results herein may be found in most elementary texts on the subject. The material in this chapter is taken primarily from [Ref. 8] and [Ref. 9], to which the reader is referred for further study.

B. CONVEX SETS

Let us return briefly to the image reconstruction problem. Consider two arbitrary functions, $f, g \in F$, the space of bounded, piecewise continuous functions on the compact set, Ω. Select some arbitrary value for a parameter, λ. We require that $\lambda \in [0,1]$. Consider the function,

$$h(\omega) = \lambda f(\omega) + (1-\lambda) g(\omega).$$

First note that as both f and g are defined on Ω, so is h. As f and g are bounded on a compact set, $M = \max\{\sup_\Omega f(\omega), \sup_\Omega g(\omega)\}$ is well defined. We know that

$$\begin{aligned} h(\omega) &\leq \lambda M + (1-\lambda)M, \text{ implying that} \\ h(\omega) &\leq M, \text{ for all } \omega\ \Omega. \end{aligned}$$

Consequently, the function, h is in F. The important items to note here are that f, g, and [illegible] $0, 1]$ were each chosen arbitrarily. We conclude, then, for any two elements $f, g \ \epsilon\ F$ and for any $\lambda\ \epsilon\ [0,1]$ the function,

$$h = \lambda f + (1-\lambda)g \quad \epsilon\ F.$$

The above example proves that the set F is a *convex set*. A set $C \subset L$, a linear vector space, is called convex if for any two elements $\mathbf{y}, \mathbf{z}\ \epsilon\ \mathrm{C}$ and $\lambda\ \epsilon\ [0,1]$,

$$\mathbf{x} = \lambda \mathbf{y} + (1-\lambda)\mathbf{z} \quad \epsilon\ \mathrm{C}.$$

Any element $\mathbf{y}\ \epsilon\ \mathrm{C}$ of the form $\mathbf{y} = \sum_{i=1}^{n} \lambda_i \mathbf{y}_i$, with $\sum_{i=1}^{n} \lambda_i = 1,\ 0 \leq \lambda_i \leq 1$ is called a *convex combination* of $\mathbf{y}_1, \mathbf{y}_2, \ldots \mathbf{y}_n$. This convex combination is called *strict* if $\lambda_i\ \epsilon\ (0,1)$ for all i. That is, the convex combination is strict if $\lambda_i \neq 0$ or 1, for all i. We now examine a fundamental characterization of convex sets.

Theorem 3. [Ref. 8] *Let C be a convex subset of L, an n-dimensional linear vector space. Every convex combination of the vectors of C is an element of C.*

Proof: For $n = 1$, the claim is trivial. Assume that the statement is true for $r \leq n-1$ where $n > 1$. Now we consider some convex combination

$$\mathbf{y} = \sum_{i=1}^{n} \lambda_i \mathbf{y}_i, \text{ where } \mathbf{y}_i \in C, \quad \sum_{i=1}^{n} \lambda_i = 1, \quad \lambda_i \geq 0.$$

If $\lambda_n = 1$, then we are done, so we suppose that $\lambda_n \neq 1$. Define

$$\lambda = \sum_{i=1}^{n-1} \lambda_i, \text{ and } \lambda_i' = \frac{\lambda_i}{\lambda}.$$

Then

$$\mathbf{y} = \lambda \sum_{i=1}^{n-1} \lambda_i' \mathbf{y}_i + \lambda_n \mathbf{y}_n.$$

Note that sum of the first term satisfies the conditions of the inductive hypothesis. That is,

$$\sum_{i=1}^{n-1} \lambda_i' = 1, \text{ and } \lambda_i' \geq 0.$$

We conclude that

$$\hat{\mathbf{y}} = \left(\sum_{i=1}^{n-1} \lambda_i' \mathbf{y}_i \right) \in C, \text{ and}$$

$$\mathbf{y} = \lambda \hat{\mathbf{y}} + \lambda_n \mathbf{y}_n.$$

Now consider the expression:

$$\begin{aligned} & \lambda + \lambda_n \\ = & \sum_{i=1}^{n-1} \lambda_i + \lambda_n \\ = & \sum_{i=1}^{n} \lambda_i \\ = & 1. \end{aligned}$$

Then by the definition of a convex set, $\mathbf{y} \in C$. □

Let $A \in \mathcal{R}^{n \times m}$, and let $\mathbf{b} \in \mathcal{R}^m$. Then it is elementary that the sets

$$\begin{aligned} G_1 &= \{\mathbf{x} : \mathrm{A}^T\mathbf{x} = \mathbf{b}\}, \text{ and} \\ G_2 &= \{\mathbf{x} : \mathrm{A}^T\mathbf{x} \geq \mathbf{b}\}, \end{aligned}$$

are convex. We prove the case of G_1.

Proof: Let $\mathbf{x}_1, \mathbf{x}_2 \in \mathrm{G}_1$. Then $\mathbf{x}_1, \mathbf{x}_2 \in \mathrm{R}^m$, and $A^T\mathbf{x}_1 = \mathrm{A}^T\mathbf{x}_2 = \mathrm{b}$. We select some value for $\lambda \in [0, 1]$, and consider:

$$\begin{aligned} & A^T(\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) \\ = \; & \lambda A^T\mathbf{x}_1 + (1-\lambda)\mathrm{A}^T\mathbf{x}_2 \\ = \; & \lambda\mathbf{b} + (1-\lambda)\mathbf{b} \\ = \; & \mathbf{b}. \end{aligned} \tag{IV.1}$$

□

One may show G_2 is convex with an identical argument. Note that the set, G_2 defines the feasible set of the linear program.

C. HYPERPLANES, POLYHEDRAL SETS, AND EXTREMA

A hyperplane H in $\mathcal{R}^n$ is a set of the form $\{\mathbf{y} : \langle\mathbf{p}, \mathbf{y}\rangle = \mathrm{k}\}$ where $\mathbf{p}$ is a nonzero vector in $\mathcal{R}^n$, and k is a given scalar. It is easily shown that the hyperplane, H, is a convex set. A hyperplane divides $\mathcal{R}^n$ into two (non-disjoint) regions, called *half-spaces*; one is defined by $\{\mathbf{y} : \langle \mathbf{p}, \mathbf{y} \rangle \ \geq \mathrm{k}\}$ and the other by $\{\mathbf{y} : \langle \mathbf{p}, \mathbf{y} \rangle \ \leq \mathrm{k}, \}$ both of which are again convex. Note that the intersection of a finite number, m, of half-spaces, called a *polyhedral set,* is also convex, since the intersection may be interpreted as $\{\mathbf{y} : \mathrm{A}^{\mathrm{T}}\mathbf{y} \geq \mathbf{b}\}$ where the i^{th} half-space is define as the set

$$\{\mathbf{y} : \langle \mathbf{a}_i, \mathbf{y} \rangle \ \geq \mathrm{b}_i\}.$$

That is, A is an $m \times n$ matrix whose columns are the vectors defining the half spaces.

To illustrate this point, we consider a simple example. Define the vectors

$$\mathbf{a}_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \mathbf{a}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \text{ and } \quad \mathbf{a}_3 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

We use the above vectors to define the three half-planes in $\mathcal{R}^2$,

$$\langle \mathbf{a}_1, \mathbf{y} \rangle \geq -2, \quad \langle \mathbf{a}_2, \mathbf{y} \rangle \geq -\tfrac{1}{4}, \text{ and } \quad \langle \mathbf{a}_3, \mathbf{y} \rangle \geq -\frac{1}{4}$$

Using the above convention for identifying the matrix A, and the vector $\mathbf{b}$, we find that

$$A = \begin{bmatrix} 0 & 1 & 1 \\ 1 & -1 & 0 \end{bmatrix}, \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} -2 \\ -\frac{1}{4} \\ \frac{1}{4} \end{bmatrix}.$$

We may identify the intersection of the half-planes as the set of vectors, $\mathbf{y}$ in $\mathcal{R}^2$ satisfying the equation,

$$A^T\mathbf{y} \geq \mathbf{b} \quad \text{or,} \quad \begin{bmatrix} 0 & -1 \\ 1 & -1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \geq \begin{bmatrix} 2 \\ -\frac{1}{4} \\ \frac{1}{4} \end{bmatrix}.$$

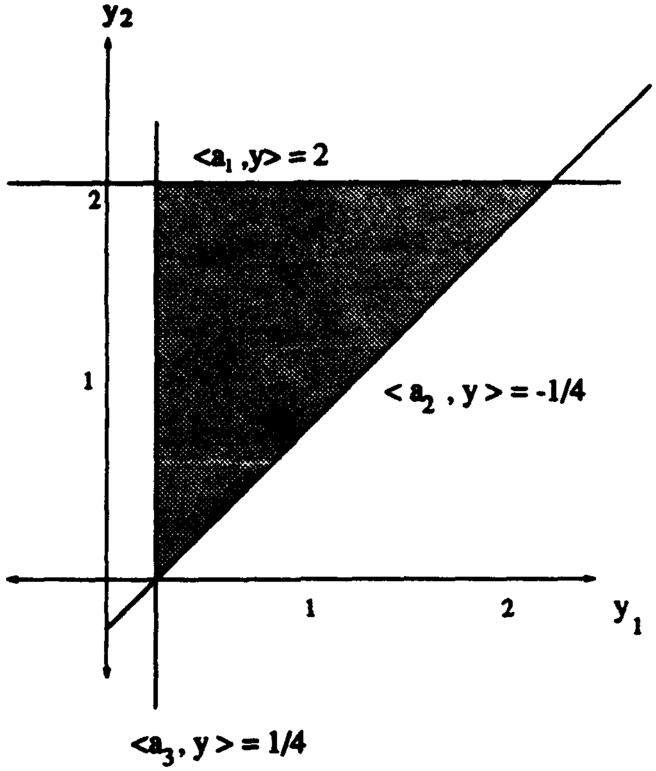

Figure 8. *The Polyhedral Set of the Example*

The intersection of these half-planes is illustrated in Figure 8.

We are interested in simplifying our optimization problem by eliminating portions of the feasible set from consideration. A critical tool in this reduction results from the notion of an *extreme point.* We here define an extreme point, and use the concept to further characterize the convex sets with which we are working in the example problems.

Let C be a convex set. We call $\mathbf{y} \in \mathrm{C}$ an *extreme point* of the set C if it can not be represented as a strict convex combination of the elements of C. Alternately,

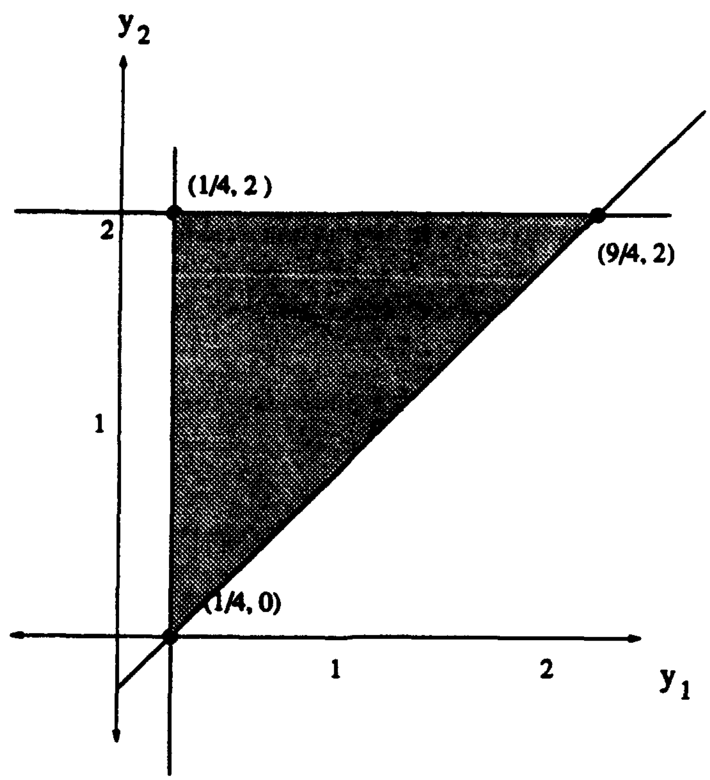

Figure 9. *Extrema of the Example Feasible Set*

the point $\mathbf{y}$ is an extreme point if, and only if, for any $\lambda \in (0, 1)$, and for any $\mathbf{x}, \mathbf{z} \in C$,

$$\mathbf{y} = \lambda\mathbf{x} + (1-\lambda)\mathbf{z} \text{ implies that}$$

$$\mathbf{y} = \mathbf{x} = \mathbf{z}.$$

Geometrically, a point $\mathbf{y}$ in a polyhedral set C is an extreme point if lies on some n linearly independent defining hyperplanes of C, where n is the rank of matrix A^T, as formed above. Two extreme points are *adjacent* if the line segment joining them is an edge of C. That is, the line segment joining them is formed by the intersection of some $n-1$ linearly independent defining hyperplanes of C. See Figure 9.

Theorem 4. [Ref. 7] *Let C be a polyhedral subset of $\mathcal{R}^n$. If C is bounded, then C has at least $n+1$ linearly independent defining hyperplanes.*

This theorem is offered without proof. However, its validity for the case of $n = 2$ is illustrated in Figure 9, where the polyhedral set in $\mathcal{R}^2$ has three independent defining hyperplanes. An immediate consequence of the above is the following:

Theorem 5. *Let C be an arbitrary bounded convex subset of $\mathcal{R}^n$. C has at least n extreme points.*

Proof: Suppose that there are fewer than n extreme points of C. Since any n linearly independent hyperplanes must intersect in a single point point in $\mathcal{R}^n$, there are fewer than $n+1$ linearly independent hyperplanes, and C is unbounded. □

D. CONVEX FUNCTIONS

We now introduce convex functions, and their primary characteristic with which we are interested. This introduction is cursory in nature. For a more detailed exploration of convexity with respect to functions, the reader is referred to [Ref. 8] and [Ref. 10].

Let $C \subset \mathcal{R}^n$ be a convex set. A function f, defined on C, is said to be *convex* if for any elements $\mathbf{x}, \mathbf{y} \in \mathrm{C}$, and $\lambda \in [0,1]$:

$$f(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \leq \lambda \mathrm{f}(\mathbf{x}) + (1-\lambda)\mathrm{f}(\mathbf{y}).$$

If f is convex, then $-f$ *is said to be concave.* Linear functions are, thus, both convex and concave. Having alluded to the utility of convex functions, we state an important result formally.

Theorem 6. [Ref. 8] *Let f be a convex function defined on a closed convex set, $C \subset \mathcal{R}^n$. Then a relative minimum of f over C is a global minimum.*

Proof: Let f have a local minimum at $\mathbf{y}_1$, and a global minimum at $\mathbf{y}_2$, with $f(\mathbf{y}_1) >$ f($\mathbf{y}_2$). Let $\lambda \in (0,1)$ be given. Because f is convex

$$f(\lambda\mathbf{y}_2 + (1-\lambda)\mathbf{y}_1) \leq \lambda f(\mathbf{y}_2) + (1-\lambda)\mathrm{f}(\mathbf{y}_1). \tag{IV.2}$$

Also, since it is assumed that $f(\mathbf{y}_1) > \mathrm{f}(\mathbf{y}_2)$, we conclude

$$\begin{aligned} \lambda f(\mathbf{y}_2) + (1-\lambda)\mathrm{f}(\mathbf{y}_1) &< \lambda f(\mathbf{y}_1) + (1-\lambda)\mathrm{f}(\mathbf{y}_1) \\ &= f(\mathbf{y}_1). \end{aligned} \tag{IV.3}$$

We now define $N_\epsilon(\mathbf{y}_1) = \{\mathbf{y} \in \mathcal{R}^n : \|\mathbf{y}-\mathbf{y}_1\| < \varepsilon\}$. That is, we define an ε neighborhood about the point, $\mathbf{y}_1$. If

$$0 < \lambda < \frac{\varepsilon}{\|\mathbf{y}_1 - \mathbf{y}_2\|}, \text{ and,}$$

$$\mathbf{y} = \lambda\mathbf{y}_2 + (1-\lambda)\mathbf{y}_1,$$

then $\mathbf{y} \in \mathrm{N}_\varepsilon(\mathbf{y}_1)$. Then

$$\begin{aligned} f(\mathbf{y}) &= f(\lambda\mathbf{y}_2 + (1-\lambda)\mathbf{y}_1) \\ &\leq \lambda f(\mathbf{y}_2) + (1-\lambda)\mathrm{f}(\mathbf{y}_1) \\ &< \lambda f(\mathbf{y}_1) + (1-\lambda)\mathrm{f}(\mathbf{y}_1) \\ &= f(\mathbf{y}_1), \end{aligned}$$

contradicting IV.3, and the fact that f has a local minimum at $\mathbf{y}_1$. We have shown, then, that only absolute minima are possible. □

If the objective function is convex (which it must be since we are considering only linear objective functions), we can be sure that we have found an optimal vector if it is locally optimal. This fact forms the basis for the Simplex algorithm, which we explore in the following chapter.

Theorem 7. *If an optimal solution to the Linear Program exists, that is, if min* $\{f(\mathbf{y})\}$ *exists and is finite for some* $\mathbf{y}$ in the feasible set, $C \subset \mathcal{R}^n$, *then there is an optimal extreme point.*

Proof: Let $\mathbf{y} \in C$ be an optimal vector, but not an extreme point. Let a linear objective function f defined on the polyhedral set C be given. Since $| f | < \infty$ at an optimal vector, one may clearly add sufficient number of hyperplanes to bound the feasible set if it is not already bounded, without changing the optimal solution. Assume that f is optimal at $\mathbf{y}$. We consider two cases.

Case 1: The vector $\mathbf{y}$ does not lie on an edge of C.

We first recognize that $\mathbf{y}$ can be written as a convex combination of the extreme points of C, since there are at least n linearly independent extreme points. Let $E = \{\mathbf{e} : \mathbf{e}$ is an extreme point of $C\}$. Let E have cardinality r. Then we may write $\mathbf{y} = \sum_{i=1}^{r} \lambda_i \mathbf{e}_i$. The linearity of the objective function, f, implies $f(\mathbf{y}) = \sum_{i=1}^{r} \lambda_i f(\mathbf{e}_i)$. Let $\mathbf{e}_j$ be some extreme point such that $f(\mathbf{e}_j) > 0$, and let us decrease the value of λ_j by $\delta > 0$ units. We may do so without leaving the feasible set since we are not on an edge. Note that if no such extreme point $\mathbf{e}_j$ exists, then we may *increase* the value of

λ_j, and the argument still holds. Call the new element of the feasible set $\mathbf{y}'$. Then

$$\begin{aligned} f(\mathbf{y}') &= f\left(\sum_{i \neq j} \lambda_i \mathbf{e}_i + (\lambda_j - \delta)\mathbf{e}_j\right), \\ &= f\left(\sum_{i \neq j} \lambda_i \mathbf{e}_i + \lambda_j \mathbf{e}_j - \delta \mathbf{e}_j)\right) \\ &= f(\mathbf{y} - \delta \mathbf{e}_j) \\ &= f(\mathbf{y}) - \delta \mathrm{f}(\mathbf{e}_j) \\ &< f(\mathbf{y}), \end{aligned}$$

implying that $\mathbf{y}$ is not the optimal vector, a contradiction. Hence if $\mathbf{y}$ is a non-extreme optimal vector, it must lie on an edge of C.

Case 2: The vector $\mathbf{y}$ lies on an edge of C.

Since $\mathbf{y}$ is not an extreme point, but is on an edge, it is on the line segment joining two extreme points, $\mathbf{e}_1$ and $\mathbf{e}_2$ of C, and may be written as $\mathbf{y} = \lambda \mathbf{e}_1 + (1-\lambda)\mathbf{e}_2$, for some $\lambda \in (0,1)$. Parameterize the line segment between the points $\mathbf{e}_1, \mathbf{e}_2$ by the equation $\mathbf{y}(\mathrm{t}) = \mathrm{t}\mathbf{e}_1 + (1-\mathrm{t})\mathbf{e}_2$, as $t : 0 \rightarrow 1$. Fix some $t \in [0,1]$, , and let $\mathbf{y}' = (1-\mathrm{t})\mathbf{e}_1 + \mathrm{t}\mathbf{e}_2$. Then

$$\begin{aligned} f(\mathbf{y}') - \mathrm{f}(\mathbf{y}) &= (1-t)f(\mathbf{e}_1) + \mathrm{t}\mathrm{f}(\mathbf{e}_2) - \lambda \mathrm{f}(\mathbf{e}_1) - (1-\lambda)\mathrm{f}(\mathbf{e}_2) \\ &= -(t-(1-\lambda))f(\mathbf{e}_1) + (\mathrm{t}-(1-\lambda))\mathrm{f}(\mathbf{e}_2) \\ &= (t-(1-\lambda))\,(-f(\mathbf{e}_1) + \mathrm{f}(\mathbf{e}_2)) \\ &\geq 0, \text{ for all } t, \text{ since } \mathbf{y} \text{ is the optimal vector.} \end{aligned}$$

Since $\mathbf{y}$ is not an extreme point, it can be represented as a strict convex combination of $\mathbf{e}_1$ and $\mathbf{e}_2$. Therefore, we may choose some $\hat{t} \in (0,1)$, such that

$\hat{t} > (1 - \lambda)$. Then

$$\begin{aligned} \hat{t} - (1 - \lambda) &> 0, \text{ implying} \\ -f(\mathbf{e}_1) + f(\mathbf{e}_2) &\geq 0. \text{ Therefore,} \\ f(\mathbf{e}_2) &\geq f(\mathbf{e}_1). \end{aligned}$$

An identical argument yields the result that $f(\mathbf{e}_2) \leq f(\mathbf{e}_1)$. We conclude that $f(\mathbf{e}_1) = f(\mathbf{e}_2)$, and that any $t \in [0, 1]$ results in an optimal vector. Choosing $t = 0$, or $t = 1$ places us at an extreme point. □

An alternate proof may be found in [Ref. 7].

E. AN ASIDE: THE CONVEX HULL

We desire to work with convex subsets of linear vector spaces, as they have useful characteristics when we attempt to solve more general optimization problems. However, there is no guarantee that an arbitrary set is convex. For such cases, we define the *convex hull* of an arbitrary set $A \subset L$, denoted Conv(A), as the set of all possible convex combinations of the elements of A, where L is a linear vector space. An example of a convex hull of a non-convex set in $\mathcal{R}^2$ is displayed in Figure 10.

Clearly, if A is convex, then Conv(A)=A. The intuitive notion that the convex hull of a set, $A \subset L$ is the smallest convex subset of L in which A is contained, and conversely, are easily proven theorems (See [Ref. 8]).

The real utility of the convex hull stems from the fact that any element of Conv(A) may be written as a convex combination of the elements of A. Generating

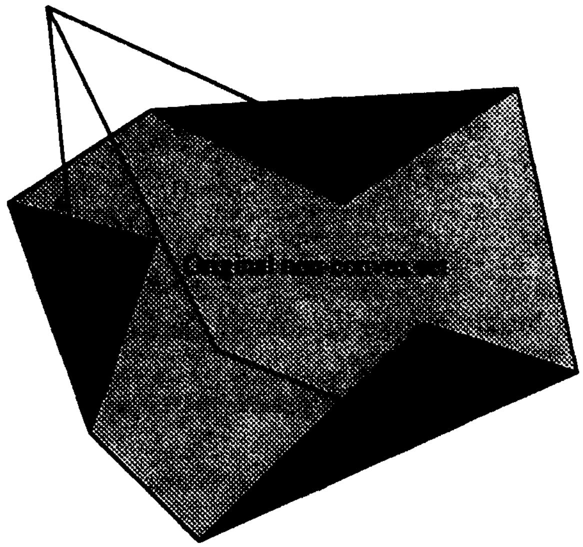

Figure 10. *Forming the Convex Hull*

the convex hull does not add any new extreme points. This is offered without proof. The interested reader should consult [Ref. 8]. Consequently, if we are solving an optimization problem with a linear objective function on a non-convex set, A, then solving it over the convex hull of the set A, rather than over the set A, itself, does not change the solution of the problem.

V. DUALITY AND THE SIMPLEX ALGORITHM

A. OVERVIEW

The concept of duality makes it possible for us to bound the optimal value for the objective function, f, and in many cases, to solve the LOP more efficiently. As before, let $\mathbf{c}$ be a vector in $\mathcal{R}^n$. Let S be an arbitrary index set. We have previously stated that for every $s \in S$, we associcate a vector $\mathbf{a}(\mathrm{s})$ in $\mathcal{R}^n$, and a scalar $b(s)$. The general form of the linear optimization problem is:

$$\begin{array}{rll} \text{minimize:} & \langle \mathbf{c}, \mathbf{y} \rangle & \\ \text{subject to} & \langle \mathbf{a}(\mathrm{s}), \mathbf{y} \rangle & \geq \; b(s), \text{ for all } s \in S \\ \text{over all} & \mathbf{y} \in \mathcal{R}^n & \end{array} \tag{V.1}$$

We know that we achieve an upper bound for the optimal value of the preference function as soon as we find an element of the feasible set. However, we have no such simple criteria for determining a lower bound. Intuitively the prospect of finding some feasible vector is less daunting than solving the problem. Using duality allows us to form an associated optimization problem, find a feasible vector in the associated problem, and use the feasible vector to derive a lower bound of the original problem. The associated optimization problem is called the *Dual.* In some cases, we may bound the original optimization from below arbitrarily well using the dual problem. We refer to the original linear optimization problem as the *primal*, P. The primal, P, and its associated dual, D, are referred to as a *Dual Pair* .

Define the *value* of a LOP to be the optimal objective function value. We seek properties that allow us to approximate the solution of a linear optimization problem arbitrarily well, and to determine when the optimal value of the linear optimization problem and its corresponding dual are the same.

This chapter, in conjunction with the previous chapter, forms the fundamental principle underlying the Simplex algorithm. The reader is again referred to [Ref. 7] and [Ref. 9] for more detailed descriptions of the material of this chapter.

B. WEAK DUALITY

We begin with the generic linear optimization problem, (V.5). The first theorem that allows us to bound the problem from below is stated here. Note that we allow for an infinite index set S.

Theorem 8. The Duality Lemma [Ref. 9] *Let the finite subset*

$$\{s_1, s_2, \ldots, s_q\} \subset S,$$

and the non-negative vector

$$\mathbf{x} = [\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_\mathrm{q}]^\mathrm{T}$$

be such that:

$$\mathbf{c} = \mathbf{a}(\mathrm{s}_1)\mathrm{x}_1 + \mathbf{a}(\mathrm{s}_2)\mathrm{x}_2 + \cdots + \mathbf{a}(\mathrm{s}_\mathrm{q})\mathrm{x}_\mathrm{q}.$$

Then for any feasible vector $\mathbf{y} = [\mathrm{y}_1, \mathrm{y}_2, \ldots, \mathrm{y}_\mathrm{n}]^\mathrm{T}$ *in the feasible set of the optimization problem, P,*

$$b(s_1)x_1 + b(s_2)x_2 + \cdots + b(s_q)x_q \leq \mathbf{c}^\mathrm{T}\mathbf{y}.$$

Proof: Since $\mathbf{y}$ is a feasible vector,

$$\langle \mathbf{a}(s_i), \mathbf{y} \rangle = \mathbf{a}(s_i)^T \mathbf{y} \geq b(s_i), \text{ for } i = 1, 2, \ldots, q.$$

Further, since $x_i \geq 0$ by assumption,

$$\sum_{i=1}^{q} b(s_i) x_i \leq \sum_{i=1}^{q} \left(\mathbf{a}(s_i)^T \mathbf{y} \right) x_i,$$

implying,

$$\begin{aligned} \sum_{i=1}^{q} b(s_i) x_i &\leq \left(\sum_{i=1}^{q} x_i \mathbf{a}(s_i) \right)^T \mathbf{y} \\ &= \mathbf{c}^T \mathbf{y} \end{aligned}$$

□

As an example, consider the problem of finding the monic second order polynomial, p_2, orthogonal to both, $p_0 = 1$, and $p_1 = x - \frac{1}{2}$. Recalling from Equation (III.2), the primal of this problem is

$$\begin{array}{lll} \text{minimize} & \frac{5}{12} + \frac{7}{12}\alpha_1 + \alpha_0 & \\ \text{subject to:} & \frac{1}{3} + \frac{1}{2}\alpha_1 + \alpha_0 & \geq 0, \\ & \frac{1}{12} + \frac{1}{12}\alpha_1 & \geq 0. \end{array} \tag{V.2}$$

Disregarding the constant in the objective function does not affect the choice of an optimal vector. Consequently, the optimal vector for (V.2) and the LOP

$$\begin{array}{llll} \text{mininize:} & \langle \mathbf{c}, \mathbf{y} \rangle & & \\ \text{subject to:} & \langle \mathbf{a}(s_1), \mathbf{y} \rangle & \geq & b(s_1) \\ & \langle \mathbf{a}(s_2), \mathbf{y} \rangle & \geq & b(s_2) \end{array} \tag{V.3}$$

where

$$\mathbf{c} = \begin{bmatrix} \frac{7}{12} \\ 1 \end{bmatrix}, \quad \mathbf{a}(\mathbf{s}_1) = \begin{bmatrix} \frac{1}{2} \\ 1 \end{bmatrix}, \quad \mathbf{a}(\mathbf{s}_2) = \begin{bmatrix} \frac{1}{12} \\ 0 \end{bmatrix}, \quad \text{and} \quad \begin{matrix} b(s_1) = -\frac{1}{3} \\ b(s_2) = -\frac{1}{12} \end{matrix} .$$

are the same.

Attempting to satisfy the hypothesis of the duality lemma, we seek a non-negative linear combination of $\mathbf{a}(\mathbf{s}_1)$ and $\mathbf{a}(\mathbf{s}_2)$ that sums to $\mathbf{c}$. That is, we seek a non-negative solution to the equation:

$$\begin{bmatrix} \frac{1}{2} & \frac{1}{12} \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \frac{7}{12} \\ 1 \end{bmatrix}.$$

Clearly, the only such vector satisfying the equation is the vector $\mathbf{x} = [1,\ 1]^T$. Consequently, the optimal value of the primal problem of V.3 can be no better than

$$\begin{aligned} & x_1\, b(s_1) + x_2\, b(s_2) \\ = \ & \left(-\tfrac{1}{3}\right) + \left(-\tfrac{1}{12}\right) = -\tfrac{5}{12}. \end{aligned}$$

Because the optimal vector of (V.2) must be the same as that of (V.3), the value of (V.2) is bounded below by

$$\frac{5}{12} - \frac{5}{12} = 0,$$

as expected.

C. THE DUAL

Having stated the duality lemma, we move to a formal definition of the dual, and similarly, the dual pair. We begin with the special case of a Linear Program. The dual of a linear program

$$\text{minimize:} \quad \langle \mathbf{c}, \mathbf{y} \rangle$$

$$\text{Subject to:} \quad A^T\mathbf{y} \geq \mathbf{b}$$

is defined to be

$$\text{maximize:} \quad \langle \mathbf{b}, \mathbf{x} \rangle$$

$$\text{Subject to:} \quad A\mathbf{x} = \mathbf{c}$$

$$\text{with} \quad x_i \geq 0, \quad \text{for } i = 1, 2, \ldots q. \tag{V.4}$$

Note that the dual of an LP is an LP itself. To be feasible above, we require a non-negative linear combination of the constraint vectors to sum to the vector **c**. The vector **f** becomes the objective vector in the dual. These facts highlight the difficulty of defining the dual of an infinite LOP. Because of the difficulty of computing infinite sums (possibly uncountably infinite), we require a variation of the dual for the infinite case.

Recall the generic LOP

$$\text{minimize:} \quad \langle \mathbf{c}, \mathbf{y} \rangle$$

$$\text{subject to} \quad \langle \mathbf{a}(s), \mathbf{y} \rangle \;\geq\; b(s), \text{ for all } s \in S$$

$$\text{over all} \quad \mathbf{y} \in \mathcal{R}^n. \tag{V.5}$$

As it proves useful in the statement of the dual, we write (V.5) in the alternate form

$$\text{minimize} \qquad \textstyle\sum_{r=1}^{n} c_r y_r$$

$$\text{subject to: } \sum_{r=1}^{n} a_r(s)y_r \geq b(s)$$

$$\text{for all} \qquad s \in S. \tag{V.6}$$

The *dual optimization problem, D*, is defined to be:

> Find a finite subset $\{s_1, s_2, \ldots, s_q\} \subset S$, and the non-negative numbers, $x_1, x_2, \ldots, x_q$, such that the expression:
>
> $$\sum_{i=1}^{q} x_i b(s_i),$$
>
> is maximized, subject to the constraints
>
> $$\begin{aligned} \sum_{i=1}^{q} x_i a_r(s_i) &= c_r, \\ \text{for } r &= 1, 2, \ldots, n. \end{aligned} \tag{V.7}$$

Tha' is, the dual of the infinite dimensional LOP is to find some optimal finite subset of the index set, and then solve the resulting LP dual. In keeping with convention, we call the process of taking a finite subset of an infinite set *discretizing*. It is important to note that the dual is, in general, a non-linear problem in $2q$ variables, since both the discretization and values for the coefficients, x_i are unknown. However, once we have chosen a subset of S, the problem is linear in the unknowns x_i. Further, one might suspect that if a sequence of discretizations of S is chosen systematically, then we may be able to arrive at an acceptable approximation of the solutic of the associated primal problem, assuming one exists. That is, we may get arbitrarily close to the solution of the dual problem, and consequently, find an arbitrarily good approximation of the solution to the infinite dimensional primal optimizaton problem by solving a sequence of Linear Programs. This is a basic premise behind solving infinite dimensional linear programs with the Simplex algorithm.

D. APPROXIMATING THE EXPONENTIAL FUNCTION

The problem of approximating the exponential function with an n^{th} degree polynomial is now analyzed more closely. Of particula: ınterest is how duality results enable us to determine the relative quality of a given approximation, and how they allow us to bound the error in the problem.

1. The Primal Problem

Recall that we stated the problem of approximating the exponential function over the inverval $[0, 3]$ as

Determine the polynomial

$$f(t) = \sum_{i=0}^{n} \alpha_i t^i$$

that mininimizes the expression

$$\sup_{t \in [0,3]} | f(t) - e^t | .$$

Let us formulate this problem in terms of the standard linear optimization problem. We relabel the index set T vice S and define it to be the interval $[0, 3]$. Realizing that the objective function above is a scalar valued function, as a first step we reformulate the problem as

$$\begin{aligned} &\text{minimize:} && \alpha_{n+1} \\ &\text{subject to:} && | \textstyle\sum_{i=0}^{n} (\alpha_i t^i) - e^t | \leq \alpha_{n+1}, \quad \text{for all } t \in T. \end{aligned}$$

Eliminating the absolute values, we replace each constraint with the equivalent pair of constraints,

$$-\sum_{i=0}^{n} \alpha_i t^i + e^t \geq -\alpha_{n+1} \quad \text{and,} \quad \sum_{i=0}^{n} \alpha_i t^i - e^t \geq -\alpha_{n+1}.$$

Rewriting, we arrive at

$$\begin{aligned} -\sum_{i=0}^{n} \alpha_i t^i + \alpha_{n+1} &\geq -e^t \quad \text{and,} \\ \sum_{i=0}^{n} \alpha_i t^i + \alpha_{n+1} &\geq e^t. \end{aligned}$$

Thus, each element of the index set T has two associated constraint vectors. Let $\mathbf{y} = [\alpha_0, \alpha_1, \ldots, \alpha_{n+1}]^{\mathrm{T}}$. We have, for each $t \in T$, a vector

$$\mathbf{a}(\mathrm{t}) = \left[\mathrm{t}^0,\ \mathrm{t}^1,\ \ldots, \mathrm{t}^n,\ 1\right]^{\mathrm{T}},$$

and the two constraints

$$\begin{aligned} -\langle \mathbf{a}(\mathrm{t}), \mathbf{y} \rangle &\geq -e^t, \text{ and} \\ \langle \mathbf{a}(\mathrm{t}), \mathbf{y} \rangle &\geq e^t. \end{aligned}$$

It proves useful in the formulation of the dual problem to distinguish the two constraints associated with each $t \in T$. As a notational device we identify the vectors

$$\mathbf{a}(\mathrm{t}^+) = \begin{bmatrix} 1 \\ t \\ t^2 \\ \vdots \\ t^n \\ 1 \end{bmatrix}, \quad \text{and} \quad \mathbf{a}(\mathrm{t}^-) = \begin{bmatrix} -1 \\ -t \\ -t^2 \\ \vdots \\ -t^n \\ 1 \end{bmatrix}.$$

It is important to note that the use of functional notation for the vectors $\mathbf{a}(\mathrm{t}^+)$ and $\mathbf{a}(\mathrm{t}^-)$ is used for convenience only. No such functional relationship exists, as there are two constraint vectors for each $t \in [0,3]$. We distinguish the vectors by labeling two sets, T^+ and T^-. Note that $T^+ = T^- = [0,3]$.

Similarly, for each $t \in T$, we have the scalar, $b(t^+) = e^t$, and $b(t^-) = -e^t$. We finally identify the objective function vector, $\mathbf{c} = [0,0,\ldots,1]^\mathrm{T} \in \mathcal{R}^{n+2}$. The final formulation of the primal problem, P, is

$$\begin{array}{lll} \text{minimize:} & \mathbf{c}^\mathrm{T}\mathbf{y} & \\ \text{subject to:} & \mathbf{a}(\mathrm{t}^+)^\mathrm{T}\mathbf{y} & \geq b(t^+) \\ & \mathbf{a}(\mathrm{t}^-)^\mathrm{T}\mathbf{y} & \geq b(t^-) \text{ for all } t \in T \\ & \text{over all} & \mathbf{y} \in \mathcal{R}^{n+2}. \end{array} \tag{V.8}$$

2. The Dual Problem

Having put the primal in the desired form, we turn our attention to the dual. Referring to the general form of the dual as in (V.7), we seek the finite subset $\hat{T} = \{t_1, t_2, \ldots, t_q\} \subset T$, and the vector, $\mathbf{x} \in \mathcal{R}^q$, that maximizes the expression

$$\sum_{i=1}^{q} x_i b(t_i)$$

while satisfying the constraints

$$\sum_{i=1}^{q} x_i a_r(t_i) \quad = \quad c_r, \qquad \text{for } r = 1, \ldots, n.$$

First make the substitutions $b(t_i) = e^{t_i}$, and $a_r(t_i) = t_i^r$. As we have defined the set T to be $T^+ \cup T^-$, the above formulation is equivelent to the following.

Find the subsets
$$\hat{T}^+ = \left\{t_1^+, t_2^+, \ldots, t_{q^+}^+\right\} \subset T^+,$$

and

$$\hat{T}^- = \left\{t_1^-, t_2^-, \ldots, t_{q^+}^-\right\} \subset T^-$$

and non-negative scalars $x_1^+, x_2^+, \ldots x_{q^+}^+$, and $x_1^-, x_2^-, \ldots x_{q^-}^-$ with which to associate each element of the respective sets, that maximizes

$$\sum_{i=1}^{q^+} e^{t_i^+} x_i^+ - \sum_{i=1}^{q^-} e^{t_i^-} x_i^-$$

and satisfies the constraints

$$\begin{aligned} \sum_{i=1}^{q^+} x_i^+ \left(t_i^+\right)^r - \sum_{i=1}^{q^-} x_i^- \left(t_i^-\right)^r &= 0, \quad \text{for } r = 0, 1, \ldots, n \\ \sum_{i=1}^{q^+} x_i^+ + \sum_{i=1}^{q^-} x_i^- &= 1. \end{aligned} \tag{V.9}$$

The formulation (V.9) may be written in the simpler form

$$\begin{aligned} &\text{maximize:} \quad \sum_{i=1}^{q} e^{t_i} x_i \\ &\text{subject to:} \quad \sum_{i=1}^{q} x_i t_i^r = 0, \quad \text{for } r = 0, 1, \ldots, n \\ &\qquad \sum_{i=1}^{q} \mid x_i \mid \leq 1. \end{aligned} \tag{V.10}$$

where $t_i \in [0, 3]$ for all i. The problems are equivalent in the respect that one may derive from a feasible solution of one a feasible solution to the other. The proof for this statement may be found in [Ref. 9].

3. Qualitative Analysis of Solutions

We begin by restating the duality lemma in the terms of the uniform approximation problem.

Theorem 9. [Ref. 9] *Let the finite subset $\hat{T} \subset T$, and the real numbers $x_1, x_2, \ldots, x_q$ be feasible for the dual problem of equation (V.10). Then the following holds for any* $\mathbf{y} \in \mathcal{R}^{n+1}$:

$$\sum_{i=1}^{q} x_i e^{t_i} \leq \sup_{t \in T} \left| \sum_{r=0}^{n} y_r t^r - e^t \right|. \tag{V.11}$$

As this is a direct consequence of the duality lemma, it is not proven here, though the proof may be found in [Ref. 9].

Let us consider the problem of approximating the exponential over T with a quadratic polynomial. Then from (V.8), the objective function vector $\mathbf{c}$ is equal to $[0,0,0,1]^T$. With each $t^+$and $t^- \epsilon\, T = [0,3]$, we associate the vectors and scalars

$$\mathrm{a}(\mathrm{t}^+) = \left[\mathrm{t}^{+^0}\ \mathrm{t}^{+^1}\ \mathrm{t}^{+^2}\ 1\right]^{\mathrm{T}} \text{ and } \mathrm{b}(\mathrm{t}^+) \ = \ \mathrm{e}^{\mathrm{t}^+},$$

and

$$\mathrm{a}(\mathrm{t}^-) = \left[-\mathrm{t}^{-^0}\ \ -\mathrm{t}^{-^1}\ \ -\mathrm{t}^{-^2}\ 1\right]^{\mathrm{T}} \text{ and } \mathrm{b}(\mathrm{t}^-) \ = \mathrm{e}^{\ \mathrm{t}^-}$$

respectively. The dual problem, from equation (V.10), is to find the set $\{\hat{t}_1, \hat{t}_2, \ldots, \hat{t}_q\} = \hat{T} \subset [0,3]$ and associated non-negative scalars that maximize

$$\sum_{i=1}^{q} x_i e^{\hat{t}_i}$$

while satisfying

$$\begin{aligned} \sum_{i=1}^{q} x_i \hat{t}_i^r &= 0, \text{ for } r = 0,1,2, \text{ and} \\ \sum_{i=1}^{q} |x_i| &\leq 1. \end{aligned} \qquad \text{(V.12)}$$

Let us arbitrarily choose the subset $\hat{T}$ to be $\{0, 1, 2, 3\}$. Hoping to apply the restated duality, Theorem 9, we first require a solution to the equation:

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 1 & 4 & 9 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}. \qquad \text{(V.13)}$$

Every such vector is of the form $\mathbf{x} = [-\alpha,\ 3\alpha,\ -3\alpha,\ \alpha]^{\mathrm{T}}$, where α is an arbitrary real number. Scaling in order to satisfy the constraint, $\sum_{i=1}^{4} \mid x_i \mid \leq 1$, we let $\mathbf{x} = \left[-\frac{1}{8},\ \frac{3}{8},\ -\frac{3}{8},\ \frac{1}{8}\right]^{\mathrm{T}}$. The hypothesis of the Theorem 9 satisfied, we conclude that the best quadratic approximation to the exponential function over $T = [0, 3]$ in the uniform norm sense, differs from e^t by at least:

$$\sum_{i=1}^{4} x_i e^{t_i} = \ -\tfrac{1}{8}e^0 + \tfrac{3}{8}e^1 - \tfrac{3}{8}e^2 + \tfrac{1}{8}e^3 \ \approx .6340.$$

E. STRONG DUALITY

Consider the three different possibilities we may encounter in the solution of the Linear Optimizaton Problem. Referring to the optimal objective function value of the minimization problem as $V(P)$, and to the optimal value of the dual as $V(D)$, we list the possible conditions, or states, of the problem as follows [Ref. 9]:

Inconsistent: (IC) The feasible set is empty, so that no solution is possible.

Bounded: (B) There exist at least one feasible vector, and among such feasible vectors, at least one is optimal.

Unbounded: (UB) There are feasible vectors such that
the objective function may be made arbitrarily small.

A *duality gap* is said to occur when $V(P) \neq V(D)$, that is, when the optimal values of the dual pair are not the same. We hope to find general conditions that preclude the existence of a duality gap. Theorems that allow us to disregard the possibility of a duality gap are called *strong duality* theorems.

1. The Dual and Convexity

We briefly characterize the dual problem as it relates to our discussion of set convexity. Before continuing, we require the definition of the *Convex Cone.* Let C be a convex subset of $\mathcal{R}^n$. The convex cone of C, denoted $\chi(C)$, is defined to be the set of all vectors $\mathbf{y} \in \mathcal{R}^n$, such that $\mathbf{y} = \lambda\mathbf{x}$, where $\lambda \geq 0$, and $\mathbf{x} \in$ C.

In Chapter IV we constructed an example of a polyhedral set using the vectors

$$\mathbf{a}(\mathrm{s}_1) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \mathbf{a}(\mathrm{s}_2) = \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \quad \text{and } (\mathrm{s}_3) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

The resultant polyhedral set is illustrated in Figure 8. The darkened region of Figure 11 illustrates the addition to the set, that together with the original polyhedral set, forms the convex cone. The darkened portions of the Figure extend to infinity.

Consider the specific case of the convex cone of the constraints of the linear optimization problem. We have expressed the constraints by $\langle \mathrm{a(s)}, \mathrm{y} \rangle \geq \mathrm{b(s)}$, for all $s \in S$. Define $A_s = \{a(s) : s \in S\} \subset \mathcal{R}^n$. We know that A_s is convex from

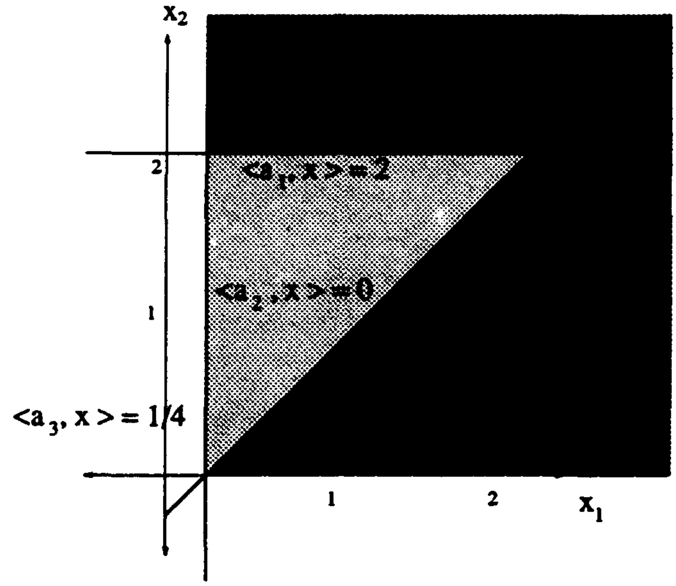

Figure 11. *Formation of the Convex Hull.*

equation (IV.1). We refer to the convex cone of A_s as the *moment cone* of the optimization problem, P, and denote $\chi(A_s)$ by M_n.

Having defined the moment cone, we arrive a fundamental characterization of the dual problem, D.

Theorem 10. [Ref. 9] *The dual problem, D, is feasible (i.e. the feasible set is not empty) if and only if* $\mathbf{c} \in \mathrm{M_n}$.

The proof may be found in [Ref. 9]. The result follows directly from the definition of the dual. An alternate interpretation of this result is as follows. The dual problem is feasible if and only if we may express the vector **c** as a non-negative combination of the constraint vectors of the linear optimization problem, P.

The following is a generalization of the theorem that allows us to express every element of the convex set, A, as a convex combination of the extreme points.

The theorem proves vital in the discussion of the Simplex algorithm, as it allows us to bound the required number of elements, $s_q \in S$, in the discretization of our index set when forming the dual.

Theorem 11. The Reduction Theorem [Ref. 9] *Let the vector* $\mathbf{z} \in \mathcal{R}^n$ *be a non-negative linear combination of the vectors,* $\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_q$. *That is,*

$$\mathbf{z} = \sum_{i=1}^{q} x_i \mathbf{z}_i,$$

with $x_i \geq 0$ *for all* i. *Then we may also write:*

$$\mathbf{z} = \sum_{i=1}^{q} x_i' \, \mathbf{z}_i, \text{ with } x_i \geq 0,$$

where at most n *of the numbers* x_i' *are nonzero. Moreover, the set of vectors* $\{\mathbf{z}_i\}$ *corresponding to the nonzero scalars* x_i' *are linearly independent.*

Proof: We first note that if $\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_q$ are already linearly independent, then $q \leq n$, and the initial representation of $\mathbf{z}$ already satisfies the theorem. Assume, then, that $q > n$, and, consequently, that the vectors, $\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_q$ are not linearly independent. Then we know that we may write

$$\sum_{i=1}^{q} \alpha_i \mathbf{z}_i = 0,$$

where at least one $\alpha_i \neq 0$. For any $r : \alpha_r \neq 0$, we have:

$$\mathbf{z}_r = -\sum_{i \neq r} \frac{\alpha_i}{\alpha_r} \mathbf{z}_i. \tag{V.14}$$

Substituting into the equation of our hypothesis, we have:

$$\mathbf{z} = \sum_{i \neq r}^{q} \left(x_i - x_r \frac{\alpha_i}{\alpha_r} \right) \mathbf{z}_i.$$

We have, then, a representation of $\mathbf{z}$ by a linear combination of $q-1$ of the vectors, $\mathbf{z}_i$. We must show, then, that the expression $(x_i - x_r\frac{\alpha-i}{\alpha_i})$ may be made non-negative, for $i = 1, 2, \ldots, r-1, r+1, \ldots, q$. Select some $\alpha_r > 0$. We can clearly do so, as if all α_i are negative, we may multiply by -1 and still have the desired result that $\sum_{i=1}^{q} -\alpha_i z_i = 0$. Then in equation (V.14), if $\alpha_i < 0$, we may conclude that

$$x_i - x_r\frac{\alpha_i}{\alpha_r} \geq 0,$$

since x_i, x_r and α_r are each nonnegative.

We now consider the case that $\alpha_i \geq 0$. Then we must show that $\frac{x_i}{\alpha_i} \geq \frac{x_r}{\alpha_r}$. We may accomplish this quite simply, by selecting the r that minimizes the expression, $\frac{x_r}{\alpha_r}$ over all $\alpha_r > 0$. We have expressed $\mathbf{z}$ as a non-negative linear combination of $q-1$ of the vectors, $\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_q$, and may continue inductively until we have the desired result. $\square$

The reduction theorem yields this immediate result. Let $\hat{S} = \{s_1, \ldots, s_q\} \subset S$, and the set of non-negative numbers $\{x_1, \ldots, x_q\}$ be feasible for the dual problem, D. That is:

$$\sum_{i=1}^{q} x_i a_r(s_i) = c_r,$$

for $r = 1, 2, \ldots, n$. Then there is a subset, $\hat{S}' = \{s_{i_1}, \ldots, s_{i_n}\}$ and a set of non-negative numbers, $\{x'_{i_1}, \ldots, x'_{i_n}\}$ that is also feasible for D. Note that we have not included the objective function of the dual in our reduction above. It is not necessarily true, then, that we need only to consider discretizations of S with cardinality n. That is, let us reduce the non-negative linear combination

$$\sum_{i=1}^{q} x_i \mathbf{a}(\mathrm{s_i}), \text{ where } \mathrm{q} > \mathrm{n}$$

to the combination

$$\sum_{i=1}^{\hat{q}} x_i' \mathbf{a}(\mathrm{s_i}),$$

where no more than n of the scalars, x' are non-zero. Then it may be that

$$\sum_{i=1}^{q} x_i b(s_i) \neq \sum_{i=1}^{\hat{q}} x_i' b(s_i).$$

Consequently, we include the optimal objective function value in the set of equations for reduction. This convention requires that we define a new moment cone, which we call, M_{n+1}.

$$M_{n+1} = \chi(A')$$

where A' is formed by the vectors,

$$\mathbf{a}'(\mathrm{s}) = \begin{bmatrix} b(s_i) \\ a_1(s_i) \\ \vdots \\ a_n(s_i) \end{bmatrix} \epsilon\, \mathcal{R}^{\mathrm{n+1}},\ \mathrm{i} = 1, 2, \ldots, \mathrm{n}.$$

The dual, then, may be stated

$$\text{maximize: } c_0$$

$$\text{subject to: } \mathbf{c} \quad \epsilon\, M_{n+1}, \text{ where } \mathrm{c} = [\mathrm{c_0}, \mathrm{c_1}, \ldots, \mathrm{c_n}]^{\mathrm{T}}$$

This formulation is useful in discussion of strong duality results.

2 Solvability Conditions

We move from the infinite case to the case of a finite index set. The following results are presented, without formal proof, though they may be found in [Ref. 9] or [Ref. 10]. These theorems enable us to determine when the dual problem has a solution. That is, we seek to determine when there exists at least one vector of our feasible set that minimizes our objective function. Note the distinction between solvability and boundedness as defined in the state section above. That is, we may have feasible vectors, but no optimal vector in our feasible set. The discussion in this section pertains to the finite case of the linear optimization problem. Readers interested in an examination of some criteria for the convergence of the *LOP* in the case of an infinite index set are referred to [Ref. 11].

Theorem 12. [Ref. 9] *Let the linear optimization problem, P, be such that M_{n+1} is closed, and the dual problem, D, is bounded. Then D has a solution.*

The proof of this theorem is straightforward. Recognize that the objective function of the dual is $f : \mathcal{R}^{n+1} \to \mathcal{R}$ by $f(z_0, z_1, \ldots, z_n) = z_0$. Then f is clearly continuous, on a compact set, and we conclude the result.

Theorem 13. [Ref. 9] *Any convex cone P defined by a finite number of vectors in $\mathcal{R}^n$ is closed, in that any convergent sequence of vectors in P converges to a vector in P.*

Coupling these observations, we conclude that any finite dual pair, (P, D), with both P and D consistent, is solveable. That is, both the primal and dual have solutions.

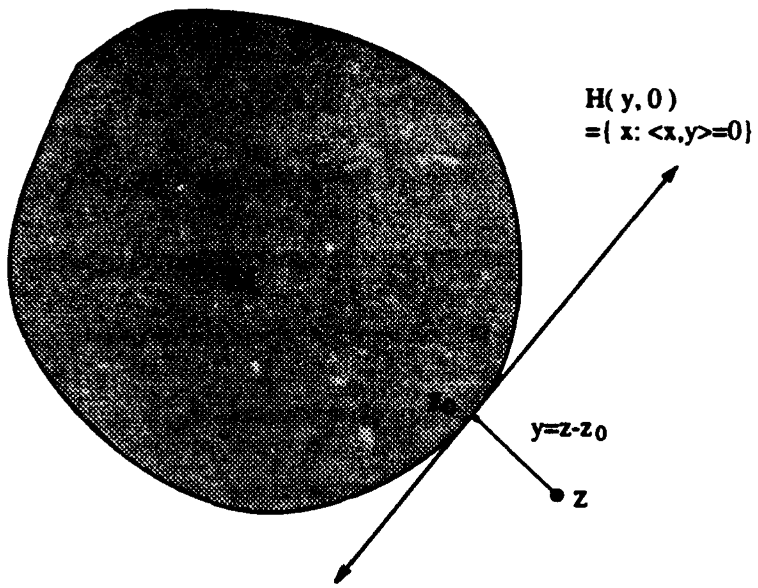

Figure 12. *The Separating Hyperplane $H(\mathbf{y}, \nu)$ of the Set, M, at the Point* $\mathbf{z}$.

3. Separating Hyperplanes

We now address the final tool that we use to eliminate a duality gap in the linear program. Let $H(\mathbf{y}, \nu) = \{\mathbf{x} \in \mathcal{R}^p : \mathbf{y}^T\mathbf{x} = \nu\}$. Then the hyperplane, $H(\mathbf{y}, \nu)$, is said to separate $\mathbf{z}$, a vector not in M, from the convex set, M, if

$$\mathbf{y}^T\mathbf{x} \leq \nu < \mathbf{y}^T\mathbf{z},$$

for all $\mathbf{x} \in$ M. Figure 12 illustrates one seperating hyperplane between the point $\mathbf{z}$ and the set M which is contained in R^2. Let $\mathbf{z}_0$ be the vector in M closest to $\mathbf{z}$ in the Euclidean norm sense. Let $\mathbf{y} = \mathbf{z} - \mathbf{z}_0$, and let $\nu = 0$. Then $H(\mathbf{y}, \nu)$ is the line orthogonal to $\mathbf{y}$ at the point $\mathbf{z}_0$.

Theorem 14. The Separation Theorem [Ref. 9] *Define $\|\mathbf{x}\|$ to be standard Euclidean 2-norm. Let $M \subset \mathcal{R}^p$ be a non-empty, closed convex set, and let $\mathbf{z}$ not be in M. Further, let $\mathbf{z}_0$ be the unique vector in M such that* $\|\mathbf{z} - \mathbf{z}_0\| \leq$

$\|\mathbf{z} - \mathbf{x}\|$ *for all* $\mathbf{x} \in M$.[1] *Finally, let* $\mathbf{y} = \mathbf{z} - \mathbf{z}_0$, *and* $\nu = (\mathbf{z} - \mathbf{z}_0)^T \mathbf{z}_0$. *Then the hyperplane,* $H(\mathbf{y}, \nu)$ *separates* $\mathbf{z}$ *from* M.

Proof: Let $\mathbf{x} \in M$, and fix $0 < \mu \leq 1$. Then

$$(1 - \mu)\mathbf{z}_0 + \mu \mathbf{x} = \mathbf{z}_0 + \mu(\mathbf{x} - \mathbf{z}_0) \in M,$$

as M is a convex set. Further,

$$\begin{aligned}
\|\mathbf{z} - \mathbf{z}_0\|^2 &\leq \|\mathbf{z} - (\mathbf{z}_0 + \mu(\mathbf{x} - \mathbf{z}_0))\|^2 \\
&= \|\mathbf{z} - \mathbf{z}_0\|^2 - 2\mu(\mathbf{z} - \mathbf{z}_0)^T(\mathbf{x} - \mathbf{z}_0) + \mu^2\|\mathbf{x} - \mathbf{z}_0\|^2,
\end{aligned}$$

which implies that

$$(\mathbf{z} - \mathbf{z}_0)^T(\mathbf{x} - \mathbf{z}_0) \leq \frac{1}{2}\mu\|\mathbf{x} - \mathbf{z}_0\|^2.$$

Let $\mu \to 0$. Then

$$(\mathbf{z} - \mathbf{z}_0)^T \mathbf{x} \leq \nu, \text{ for any } \mathbf{x} \in M.$$

Then by the definition of a separating hyperplane, we have only to show that $\nu < \mathbf{y}^T \mathbf{z}$. Since $\mathbf{z}$ is not in M,

$$0 < \|\mathbf{z} - \mathbf{z}_0\|^2 = (\mathbf{z} - \mathbf{z}_0)^T(\mathbf{z} - \mathbf{z}_0)$$

$$= \mathbf{y}^T\mathbf{z} - \mathbf{y}^T\mathbf{z}_0 = \mathbf{y}^T\mathbf{z} - \nu.$$

□

The separating hyperplane defined above is a necessary tool in the elimination of duality gaps in the finite linear optimization problem.

[1] That such a *unique* vector exists is proven in [Ref. 9].

4. The Strong Duality Theorem

We close this section with a statement and proof of a fundamental theorem of linear optimization, which states sufficient conditions for the absence of a duality gap in the dual pair, (P, D).

Theorem 15. [Ref. 8] *Let the dual pair, (P, D) satisfy the following assumtions.*

1. *The dual problem is consistent and has a finite value $V(D)$.*
2. *The moment cone, M_{n+1} is closed.*

Then (P) is consistent, and $V(P) = V(D)$. That is, no duality gap occurs.

Proof: Let let the vector, $\mathbf{c} = [c_0, c_1, \ldots, c_n]^T \in M_{n+1}$, be an optimal solution of the dual problem. Then, for any $\varepsilon > 0$, the vector, $\mathbf{c}' = [c_0 + \varepsilon, c_1, \ldots, c_n]^T$ is not in M_{n+1}. As we are assuming that M_{n+1} is closed, we conclude that there is a hyperplane separating the vector $\mathbf{c}'$ from M_{n+1}. Consequently, there exists some vector $\mathbf{y} = [y_0, y_1, \ldots, y_n]^T \in \mathcal{R}^{n+1}$, with $\mathbf{y} \neq 0$, such that

$$\sum_{r=0}^{n} x_r y_r \ \leq \ 0 \ < \ y_0(c_0 + \varepsilon) + \sum_{r=1}^{n} c_r y_r,$$

for $\mathbf{x} = [x_0, x_1, \ldots, x_n]^T \in M_{n+1}$. Let $\mathbf{x} = \mathbf{c}$. Then $y_0\ \varepsilon > 0$, implying $y_0 > 0$. Now let,

$$\mathbf{x} = [b(s), a_1(s), \ldots, a_n(s)]^T \ \in A'_s \ \subset \ M_{n+1}.$$

where $s \in S$, and $a_r(s)$ is the r^{th} component of the constraint vector associated with s. Then

$$\sum_{r=1}^{n} a_r(s) \left(\frac{-y_r}{y_0} \right) \geq b(s),$$

which implies,

$$y' = \left[\frac{-y_1}{y_0}, \ldots, \frac{-y_n}{y_0}\right]^T \epsilon\, \mathcal{R}^n$$

is feasible for the primal, P. Further,

$$0 < y_0(c_0 + \varepsilon) + \sum_{r=1}^{n} c_r y_r,$$

$$\Rightarrow \sum_{r=1}^{n} c_r \left(\frac{-y_r}{y_0}\right) < c_0 + \varepsilon.$$

Applying the duality lemma, we conclude that

$$V(P) \;\leq\; \sum_{r=1}^{n} c_r y'_r \;<\; c_0 + \varepsilon \;=\; V(D) + \varepsilon \;\leq\; V(P) + \varepsilon,$$

implying

$$V(P) - \varepsilon \leq V(D) \leq V(P),$$

for *any* ε. □

Of a final note, if the index set of our constraints is finite, then we may conclude immediately that no duality gap exists in the dual pair, (P, D). This follows directly from the above theorem in conjunction with Theorems 12 and 13.

F. THE SIMPLEX ALGORITHM

We present a very brief introduction to the Simplex algorithm, and use it to solve a simple LOP. This section is not intended to illustrate the implementation of the algorithm in any specific form. Rather, this section attempts to explain the algorithm as it exploits the results of the duality concepts above. The problem is assumed to be infinite-dimensional in this presentation.

We begin with a problem, P, of the form:

$$\text{Minimize:} \quad \sum_{r=1}^{n} c_r y_r$$

$$\text{subject to:} \quad \sum_{r=1}^{n} a_r(s) y_r \;\geq\; b(s), \text{ for all } s \in S.$$

Then we write the dual, D:

$$\text{Maximize:} \quad \sum_{i=1}^{n} b(s_i) x_i$$

$$\text{subject to:} \quad \sum_{i=1}^{n} a_r(s_i) x_i \;= c_r, r = 1, 2, \ldots, n$$

$$s_i \in S, \qquad x_i \geq 0.$$

Choose some subset, $\sigma = \{s_1, s_2, \ldots, s_n\} \subset S$, and a vector $\mathbf{x} = [\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_n]^T$ that is feasible for the dual. The methods to arrive at an initial feasible vector, provided one exists, may be found in any Linear Program text. In particular, the reader is referred to [Ref. 7]. We derive a vector $\mathbf{y}$ from our choice of σ, which is associated with the primal problem. As a matter of convenience, we abbreviate this set of values, $\{\sigma, \mathrm{x}, \mathbf{y}\}$. We require that the vectors, $\mathbf{a}(\mathrm{s_i})$ be linearly independent. That we may always find a set of linearly independent vectors is assumed in this presentation.

Forming our matrix A as before, we know that the linear independence of the vectors ensures that there is a unique vector, $\mathbf{x}$ satisfying:

$$A\mathbf{x} = \mathbf{c},$$

since we are feasible in the dual. Define the *discretized primal* to be the linear program that results in considering only the finite subset of the index set S. Let

$A(s_1, s_2, \ldots, s_n) = [\mathbf{a}(s_1), \mathbf{a}(s_2), \ldots, \mathbf{a}(s_n)]$, with $\mathbf{b}(s_1, \ldots, s_n)$ defined in the same manner. From the discretization, σ, we look for a vector, $\mathbf{y}$, that is feasible for the discretized primal, P. We note that one such vector, $\mathbf{y}$ solves the equation:

$$A^T(s_1, s_2, \ldots, s_n)\mathbf{y} = \mathbf{b}(s_1, s_2, \ldots, s_n).$$

Then

$$\mathbf{y} = (\mathrm{A}^{\mathrm{T}})^{-1}\mathbf{b}.$$

The set of values of σ and the vector $\mathbf{y}$ that is formed in the manner above is called a *basic solution* of the LOP. The steps of the algorithm, to this point are:

1. Select a subset, $\sigma \subset S$, such that the vectors, $\mathbf{a}(s_1), \mathbf{a}(s_2), \ldots, \mathbf{a}(s_n)$ are linearly independent.
2. Compute the unique non-negative solution to the equation, $A\mathbf{x} = \mathbf{c}$.
3. Compute the solution to the system, $A^T\mathbf{y} = \mathbf{b}$, for the discretized primal.

Return to the problem of approximating the exponential with a quadratic polynomial over the interval $[0, 3]$. We have formulated the problem with the constraint vectors of the index sets, $\mathbf{a}(t^+)$, and $\mathbf{a}(t^-)$, given by

$$\mathbf{a}(t^+) = \begin{bmatrix} 1 \\ t \\ t^2 \\ 1 \end{bmatrix}, \quad \text{and} \quad \mathbf{a}(t^-) = \begin{bmatrix} -1 \\ -t \\ -t^2 \\ 1 \end{bmatrix}.$$

Additionally, the constraint scalars were defined to be $b(t^+) = e^t$, and $b(t^-) = -e^t$, and the objective function vector was given by $\mathbf{c} = [0, 0, 0, 1]^{\mathrm{T}}$. The problem is

$$\begin{aligned} \text{minimize:} \quad & \mathbf{c}^T\mathbf{y} \\ \text{subject to:} \quad & \mathbf{a}(\mathbf{t}^+)^T\mathbf{y} \geq b(t^+) \\ & \mathbf{a}(\mathbf{t}^-)^T\mathbf{y} \geq b(t^-) \text{ for all } t \in T \\ \text{over all} \quad & \mathbf{y} \in \mathcal{R}^4. \end{aligned}$$

Step One: Arbitrarily select σ to be composed of the union of the sets $\sigma_1 = \{0,2\} \subset T^-$ and $\sigma_2 = \{1,3\} \subset T^+$.

Step Two: Compute the solution of the system

$$\begin{bmatrix} -1 & 1 & -1 & 1 \\ 0 & 1 & -2 & 3 \\ 0 & 1 & -4 & 9 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}. \tag{V.15}$$

The solution of this system is given by

$$\mathbf{x} = \begin{bmatrix} \frac{1}{8} & \frac{3}{8} & \frac{3}{8} & \frac{1}{8} \end{bmatrix}.$$

Step Three: Compute the solution of

$$\begin{bmatrix} -1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 \\ -1 & -2 & -4 & 1 \\ 1 & 3 & 9 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ k \end{bmatrix} = \begin{bmatrix} -e^0 \\ e^1 \\ -e^2 \\ e^3 \end{bmatrix}. \tag{V.16}$$

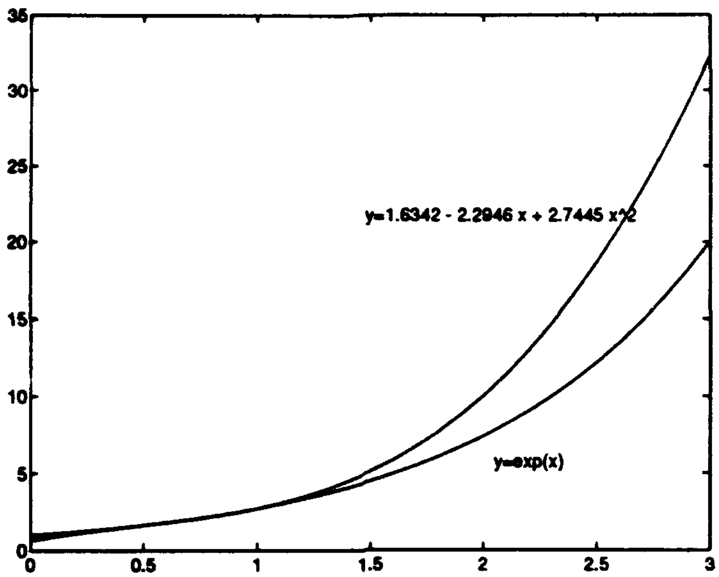

Figure 13. *A First Approximation of the Exponential Function.*

The vector $\mathbf{y} = [1.6342 \;\; -2.2946 \;\; 2.7445 \;\; .6342]^{\mathrm{T}}$ is the unique solution of this system. That is, $\mathbf{y}$ is feasible for the discretized primal. The first approximation is given by

$$p_2(x) \;=\; 1.6342 - 2.2946x + 2.7445x^2. \qquad \text{(V.17)}$$

The graph of the exponential versus the approximation is given in Figure 13.

We here introduce a lemma that offers us a termination criteria for the algorithm.

Theorem 16. The Complementary Slackness Theorem [Ref. 9] *Let the set, $\{\sigma, \mathbf{x}, \mathbf{y}\}$ be as above. If the vector $\mathbf{y}$ is feasible for the non-discretized primal P, and the following holds:*

$$x_i\left(\sum_{r=1}^{n} a_r(s_i)y_r - b(s_i)\right) = 0, \; \textit{for } r = 1, 2, \ldots, n.$$

Then we may conclude, that if the vector, $\mathbf{y}$, as determined in step 3, is feasible for the primal, P, we have found the optimal vector in our problem.

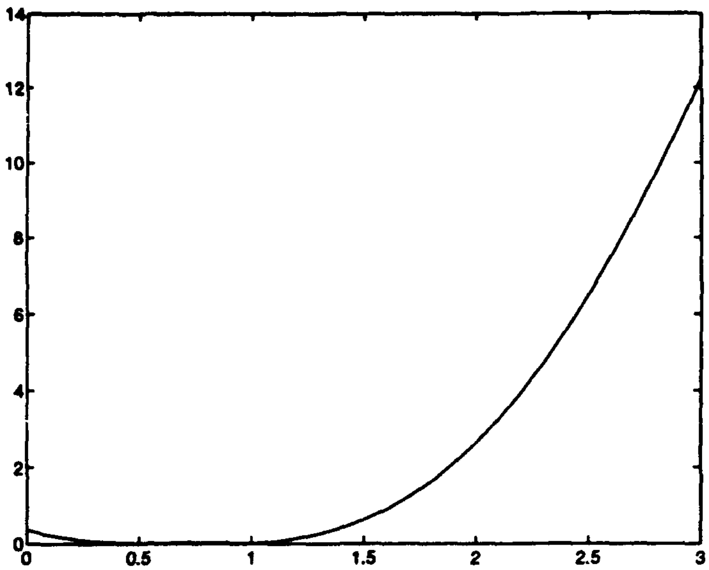

Figure 14. *Absolute Error in the First Exponential Approximation.*

In the current approximation problem, we find that the current solution does not satisfy this criteria. We observe the graph of the absolute difference between the functions and find that the error exceeds .6342 over the latter portion of the interval. See Figure 14.

The remainder of the algorithm is a sequence of *exchange steps* that replace existing elements of the set, σ, with elements that improve the value of the dual problem, D, and consequently, improve the bound of $V(P)$. The method of selecting new elements to the set, σ, may change with implementation, but it should be noted that exactly one element of the set σ is replaced at a given step, in any implementation. Recalling from our discussion of extreme points of our feasible set, that strategy ensures that the algorithm looks to adjacent extreme points for optimality.

We conduct one such exchange. Note that the error is most severe at the

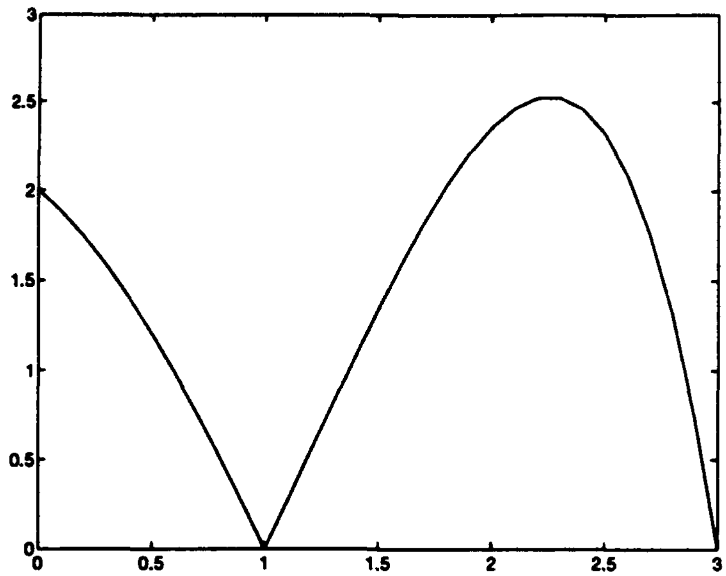

Figure 15. *Absolute Error in the Second Exponential Approximation.*

point $t = 3$. Then it is logical to seek a better solution at that point. Then we let $\sigma_1 = \{0, 3\}$. The new system of equations requiring a solution in step 3 is

$$\begin{bmatrix} -1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 \\ -1 & -3 & -9 & 1 \\ 1 & 3 & 9 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ k \end{bmatrix} = \begin{bmatrix} -e^0 \\ e^1 \\ -e^3 \\ e^3 \end{bmatrix}. \tag{V.18}$$

The solution of the above system is given by

$$\mathbf{y} = \left[0, 0, \frac{1}{2}, \frac{1}{2}\right]^T.$$

We find that the error is decreased. The absolute error is given in Figure 15.

We note that the solution is not feasible for the entire interval, since there exist points where the error exceeds .5. Thus, we would look to adjacent extreme point solutions and repeat the process until we arrive at a discretized solution that is feasible throughout the interval.

VI. RECONSTRUCTION FORMULATION AND SOLUTION

A. OVERVIEW

Having laid the complete foundation, we formulate the image reconstruction optimization problem. The first portion of this chapter addresses the conceptual aspects of the problem, while in the latter portion we use the Simplex algorithm to solve a simple reconstruction problem. We conclude the chapter with a brief discussion of the merits and drawbacks of a Linear Programming approach to the reconstructi(problem.

B. TARGET FUNCTIONS AND NORMS

The problem we wish to solve is to find the density function, f, that produces the observed sampled Radon transform. As the problem is ill-posed, we must define some preference function by which to compare the quality of the infinitely many density functions that satisfy the above requirement. We do so by specifying some function, g, defined over Ω, which is assumed to represent the most likely density of the image. That is, of all density functions that produce the observed transform data, we seek that which is most like what we expect to find. How we determine the function g is not a matter of discussion here. We only assume that we know some such function.

The problem of how to determine the best solution becomes one of finding the density function that produces the observed transform that is "closest" to g in some sense. We choose the infinity norm, or max norm, to measure closeness. Let P be a polygonal partition of the compact set $\Omega \subset \mathcal{R}^2$, consisting of the n polygons, $s_1, s_2, \ldots, s_n$. Recall that the function $\psi_j(\omega)$ is defined to be the characteristic function of polygon j in P. Imposing the restriction that the optimum density be constant over each polygon, the density takes on the form,

$$f(\omega) = \sum_{j=1}^{n} \alpha_j \psi_j(\omega).$$

We seek a density, f defined over Ω that minimizes the following:

$$\|f(\omega) - g(\omega)\|_\infty \;=\; \max_{j=1,2,\ldots,n} \left\{ \max_{\omega \epsilon s_j} \mid \alpha_j - g(\omega) \mid \right\}. \tag{VI.1}$$

We also choose some $\hat{\varepsilon} > 0$ and insist that

$$\begin{aligned} \tilde{f}_P - \mathbf{b} &\leq \hat{\varepsilon}, \text{ and} \\ f &\geq 0, \end{aligned}$$

where the vector inequality is componentwise. Recall that $\tilde{f}_P$ is defined to be the sampled transform of the density f for partition, P. The vector $\mathbf{b}$ is the observed sample Radon transform. The non-negativity constraint stems from the physical nature of the problem. That is, we do not accept solutions that attribute negative density to physical objects.

Before continuing, let us consider the objective function of equation (VI.1). Recall that our attention is fixed on density functions defined on Ω, a compact subset of $\mathcal{R}^2$. Let us first fix our attention on some polygon, s_j, in the polygonal partition P. Let M_j denote the largest absolute difference between our target function, g and the scalar, α_j, that we associate with the polygon, s_j. That is, $f(\omega) = \alpha_j$, for all $\omega \epsilon s_j$. The term

$$\max_{\omega \epsilon s_j} \mid \alpha_j - g(\omega) \mid$$

is well defined, as both functions are piecewise continuous over the compact set, s_j. The objective function is defined to be the largest of the M_j values over all polygons.

As the problem is not linear, we write an equivalent formulation:

$$\begin{aligned}
\text{minimize:} \quad & k \\
\text{subject to:} \quad \|\tilde{f}_P - \mathbf{b}\|_\infty & \leq \hat{\varepsilon} \\
\alpha_j + k & \geq g(\omega)\psi_j(\omega), \quad \text{for all } \omega \epsilon \Omega, \\
-\alpha_j + k & \geq -g(\omega)\psi_j(\omega), \quad \text{for all } \omega \epsilon \Omega, \\
\alpha_j & \geq 0, \quad \text{for all } j.
\end{aligned} \tag{VI.2}$$

Suppose the target function, g, is chosen to be continuous over Ω, and further suppose that $\tilde{g}_P = \mathbf{b}$, where $\tilde{g}_P$ is the sample transform of the target density, g. If the method is to prove worthwhile, we expect that the test density function is optimal. That is, if the test and target densities are the same, we can expect to find an arbitrarily good approximation of the test density. We state the above formally.

Theorem 17. *Let g be a non-negative, continuous function defined on the set Ω. Additionally, let values for $\varepsilon > 0$ and $\hat{\varepsilon} > 0$ be given. Then there exists*

some partition, P, of n polygons, and an associated function, $f = \sum_{j=1}^{n} \alpha_j \psi_j$, so that the optimum value of the linear optimization problem:

$$
\begin{array}{llll}
\textit{minimize:} & k & & \\
\textit{subject to:} & \|\tilde{f}_P - \tilde{g}_P\|_\infty & \leq \ \hat{\varepsilon} & \text{(VI.3)} \\
 & \alpha_j + k & \geq \ g(\omega)\psi_j, \quad \textit{for all } \omega \in \Omega, & \text{(VI.4)} \\
 & -\alpha_j + k & \geq \ -g(\omega)\psi_j, \quad \textit{for all } \omega \in \Omega, & \text{(VI.5)} \\
 & \alpha_j & \geq 0, \quad \textit{for all } j & \text{(VI.6)}
\end{array}
$$

is less than ε.

Proof: We show that we may find a feasible vector for any value of k, and consequently, for $k < \varepsilon$. The proof depends on the continuity of the sample Radon transform. That is, let g be any continuous function defined on Ω, and let $\hat{\varepsilon} > 0$ be given. Then we wish to show that there exists some $\delta_1 > 0$ and some partition P_{δ_1} such that the following property holds:

$$\|f - g\|_\infty < \delta_1 \Rightarrow \|\tilde{f}_{P_{\delta_1}} - \tilde{g}_{P_{\delta_1}}\|_\infty < \hat{\varepsilon}.$$

Let $h(\omega) = | f(\omega) - g(\omega) |$. Recall that for a fixed partition, a single integral over a strip, q, defining the sample transform takes on the form

$$\tilde{h}_q = \sum_{i=1}^{n} \left(\int_{s_i} h(\omega) \ \gamma_q(\omega) d\omega \right),$$

where $\gamma_q(\omega)$ is the characteristic function of the q^{th} strip. Let M denote the area of the largest polygon in our partition, choose our δ_1 to be less than $\frac{\hat{\varepsilon}}{Mn}$. That is, let the functions f and g differ by no more than δ_1 in the uniform norm sense. We have already proven that we may do so for some partition. Then we know for each element of our sample data vector

$$
\begin{aligned}
\tilde{h}_q &= \sum_{i=1}^{n} \left(\int_{s_i} h(\omega)\ \gamma_q(\omega) d\omega \right), \\
&\leq \sum_{i=1}^{n} \left(\int_{s_i} \frac{\hat{\varepsilon}}{Mn}\ \gamma_q(\omega) d\omega \right), \\
&\leq \sum_{i=1}^{n} \left(\frac{\hat{\varepsilon}}{Mn} M \right), \\
&\leq \hat{\varepsilon}. \qquad \text{(VI.7)}
\end{aligned}
$$

As VI.7 holds for each of the finite number of sample integrals, we may conclude that

$$\|\tilde{f}_{P_{\delta_1}} - \tilde{g}_{P_{\delta_1}}\|_\infty < \hat{\varepsilon}.$$

Thus, if we can disregard constraints, VI.4, VI.5, and VI.6, for any $\hat{\varepsilon} > 0$, we may find a partition P_{δ_1} that ensures

$$\|f - g\|_\infty \ \leq \ \delta_1$$

so that

$$\|\tilde{f}_{P_{\delta_1}} - \tilde{g}_{P_{\delta_1}}\|_\infty \ \leq \ \hat{\varepsilon}.$$

This implies that for any value of $\hat{\varepsilon}$, constraint (VI.3) is met.

Temporarily disregarding the constraint $\|\tilde{f}_P - \tilde{g}_P\|_\infty \leq \hat{\varepsilon}$, we have the less restrictive optimization problem:

$$
\begin{aligned}
\text{mininize:} \quad & k \\
\text{subject to:} \quad & \alpha_j + k \geq g(\omega)\psi_j(\omega), \ \text{for all } \omega \in \Omega, \\
& -\alpha_j + k \geq -g(\omega)\psi_j(\omega), \ \text{for all } \omega \in \Omega, \\
& \alpha_j \geq 0, \ \text{for all } j. \qquad \text{(VI.8)}
\end{aligned}
$$

That the value of the above optimization problem may be made arbitrarily small is a direct result of the fact that we may represent any continuous function, g on Ω arbitrarily well by a function of the desired form, per Theorem 2. Denote the partition for which the value of k is less than ε by P_{δ_2}. The term δ_2 is the largest largest difference between two points in the same polygon of P. That is,

$$x, y \in s_j \in P_{\delta_2} \;\Rightarrow\; \|x - y\|_\infty < \delta_2.$$

Finally, choosing δ to be the $\min\{\delta_1, \delta_2\}$, constraint VI.3 is met, as are constraints VI.4 through VI.6 for $k < \varepsilon$. Then the problem is feasible. As this is a finite dimensional problem, we employ Theorem 13 to ensure that a solution to the problem exists. Therefore, the optimal value of the optimization problem is less than ε. □

From the above claim, we expect that if our partition is sufficiently fine, then we may reasonably expect to find an acceptable approximation to the solution of our optimization problem.

C. PROBLEM STANDARDIZATION

We wish to understand the above formulation as it relates to our definition of the general linear optimization problem. Before proceeding, it is vital to note that we are formulating the problem *after* we have generated a polygonal partition, P, of Ω. Throughout this section, we assume that P contains the n polygons, $s_1, \ldots, s_n$.

1. Inner Product Constraints: Refining the Feasible Set

Before considering the constraints themselves, recall that the polygonal partition, P, forms an n-dimensional basis for a subset of the space of functions from which we select our optimal function. We may, consequently, think of any density function f as a vector $\mathbf{y} \in \mathcal{R}^n$, where the j^{th} component of $\mathbf{y}$ is the scalar value of the density on polygon, s_j. For reasons that become clear shortly, we augment the vector of decision variables to be $\mathbf{y} = [\alpha_1,\ \alpha_2, \ldots, \alpha_n,\ \mathrm{k}]^T \in \mathcal{R}^{n+1}$.

We divide our constraints into three distinct classes:

1) Strip based constraints,
2) Polygon based constraints, and
3) Ω based constraints.

First consider the strip based constraints. We require that the sample transform of the optimal objective density be within a specified tolerance of the observed sample transform. The constraints were identified in the previous section by the equation

$$\|\tilde{f}_P - \mathbf{b}\|_\infty \ \leq \ \hat{\varepsilon}.$$

Eliminating the norm above results in the two constraints

$$-\tilde{f}_P \ \geq \ -\mathbf{b} - \hat{\varepsilon}, \tag{VI.9}$$

and

$$\tilde{f}_P \ \geq \ \mathbf{b} - \hat{\varepsilon}. \tag{VI.10}$$

Let m denote the number of strips used to generate the partition, P. Let $Q = \{q_1, q_2, \ldots, q_m\}$ be the set of such strips. Then for each $q \in Q$, we require that the j^{th} component of our constraint vector be determined by the following rule:

$$a_j(q_i) = \text{area}(s_j) \max_{\omega \in \Omega} \{\gamma_q(\omega)\ \psi_j(\omega)\}.$$

Of course, this convention is the same as that of Chapter III. The j^{th} component of the vector, $\mathbf{a}(q_i)$ is the area of the j^{th} polygon if the polygon falls within strip q_i, and zero otherwise. As before, we wish to consider the two constraints associated with each strip separately, and define Q^- and Q^+ to index constraints (VI.9) and (VI.10) respectively.

The right hand side of each constraint is also determined as in Chapter III. That is, $b(q-)$ and $b(q_+)$ as in equations (VI.9) and (VI.10), where b_i is the data from strip q_i of our sample transform. We append a zero to each strip based constraint vector, as each is independent of the value k.

The polygon based constraints are found entirely in the requirement that our density function be non-negative. In the initial formulation, the requirement was written

$$\alpha_j \geq 0, \quad \text{for all } j.$$

That is, we require that the density assigned to each polygon in the optimal vector be non-negative. Let $P = \{s_1, s_2, \ldots, s_n\}$ be the fixed partition. Then the constraint vector associated with each polygon is form[illegible] by the simple rule:

$$
\begin{aligned}
a_j(s_i) &= \max_{\omega\epsilon\Omega}\{\psi_j(\omega)\ \psi_i(\omega)\}, \\
&= \begin{cases} 1, & \text{if } i = j \\ 0, & \text{otherwise,} \end{cases}
\end{aligned}
$$

for $i, j = 1, 2, \ldots, n$.

We append a zero to each $\mathbf{a}(\mathrm{s_i})$ as in the case of the strip based constraints, as we are selecting a vector from $\mathcal{R}^{n+1}$. Clearly, $b(s_i) = 0$, for all i. Then the vector form of each polygon based constraint is:

$$\langle \mathbf{a}(\mathrm{s_i}), \mathbf{y} \rangle \ \geq 0, \ \text{for } \mathrm{i} = 1, 2, \ldots, \mathrm{n}.$$

As our third class of constraints corresponds to the set Ω, we may correctly infer that the final index sets are infinite. These index sets provide the constraints that facilitate a comparison of solution quality. There are, in fact, two such index sets, Ω^+ and Ω^-, as we again eliminate the explicit use of the infinity norm from the formulation. In the initial problem statement, these constraints were written

$$
\begin{aligned}
\alpha_j + k &\geq g(\omega)\ \psi_j(\omega), \\
-\alpha_j + k &\geq \ -g(\omega)\ \psi_j(\omega), \quad \text{for all } \omega \ \epsilon\ \Omega.
\end{aligned}
$$

We focus only on the former, as Ω^- is formulated in a nearly identical manner, and the process has been executed in the strip based constraints. We desire constraints of the form $\langle \mathbf{a}(\omega^+), \mathbf{y}\rangle \geq \mathrm{b}(\omega^+)$, for all $\omega^+ \epsilon\, \Omega^+$. Let

$$\begin{aligned} a_j(\omega^+) &= \psi_j(\omega^+), \quad \text{for } j = 1, 2, \ldots, n,\ \omega^+ \epsilon\, \Omega^+, \\ a_{n+1}(\omega_i^+) &= 1, \qquad \text{for all } \omega^+ \epsilon\, \Omega^+. \end{aligned}$$

The associated $b(\omega^+)$ is defined to be $g(\omega^+)$. The constraints associated with the set Ω^- are formed in exactly the same manner, with sign changes as appropriate.

Concluding, we define our index set $T = Q^+ \cup Q^- \cup P \cup \Omega^+ \cup \Omega^-$.

D. THE IMAGE RECONSTRUCTION DUAL

The image reconstruction optimization problem, as we have formed it, is a specific example of the uniform approximation problem. Consequently, we find some strong similarities in its dual problem to the dual of the approximation of the exponential function. Let us derive the dual, D, of our image reconstruction problem. Note that this section is included in the interest of completeness. The material herein is complicated and is not especially enlightening. The reader may wish to skip this section.

We seek a subset, $\{t_1, t_2, \ldots, t_q\} \subset T$, and the non-negative vector $\mathbf{x} = [\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_q]^{\mathrm{T}}$ that maximize the equation

$$\langle \mathbf{b}, \mathbf{x}\rangle$$

and satisfies

$$\langle \mathbf{x}, \mathbf{a}(\mathbf{t}) \rangle = \mathbf{c}.$$

We address the selection of the subset first. Consider the strip-based constraints, associated with $Q \subset T$. Recall that $Q = Q^+ \cup Q^-$. We seek some subset of each of these sets. We denote these subsets $\hat{Q}^+$, and $\hat{Q}^-$. With each element of each subset, we associate a non-negative real number, $x(\hat{q}_j^+)$, for $j = 1, \ldots, n_{\hat{q}^+}$, and $x(\hat{q}_j^-)$, for $j = 1, \ldots, n_{\hat{q}^-}$.

Considering the index set, P, with which we associated the polygon based constraints, we seek some non-negative value $x(\hat{s}_j)$ to associate with each constraint of a subset $\hat{P} \subset P$. Following the above convention, we let $j = 1, \ldots, n_{\hat{P}}$.

Let us move to the infinite index sets, Ω^+ and Ω^-. As we noted above, there are two classes of constraint vectors associated with our index sets, Ω^+ and Ω^-. In particular,

$$\begin{bmatrix} \psi_1(\omega^+) \\ \psi_2(\omega^+) \\ \vdots \\ \psi_n(\omega^+) \\ 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} -\psi_1(\omega^-) \\ -\psi_2(\omega^-) \\ \vdots \\ -\psi_n(\omega^-) \\ 1 \end{bmatrix}.$$

For each of our index sets, Ω^+ and Ω^-, we seek some discretization

$$\hat{\Omega}^+ = \{\omega_1^+, \omega_2^+, \ldots, \omega_{n_{\hat{\Omega}^+}}^+\}, \text{ and}$$

$$\hat{\Omega}^- = \{\omega_1^-, \omega_2^-, \ldots, \omega_{n_{\hat{\Omega}^-}}^-\}, \text{ as well as non-negative scalars,}$$

$$x(\omega_1^+), x(\omega_2^+), \ldots, x(\omega_{n_{\hat{\Omega}+}}^+), \text{ and}$$

$$x(\omega_1^-), x(\omega_2^-), \ldots, x(\omega_{n_{\hat{\Omega}-}}^-).$$

Then the dual D is to find the above discretization and non-negative x values that maximize the expression:

$$\sum_{i=1}^{n_{\hat{Q}+}} x(q_i^+)b(q_i^+) - \sum_{i=1}^{n_{\hat{Q}-}} x(q_i^-)b(q_i^-) + \sum_{i=1}^{n_{\hat{P}}} x(s_i)b(s_i) + \sum_{i=1}^{n_{\hat{\Omega}+}} x(\omega_i^+)b(\omega_i^+) - \sum_{i=1}^{n_{\hat{\Omega}-}} x(\omega_i^-)b(\omega_i^-),$$

while satisfying the constraints:

$$\sum_{i=1}^{n_{\hat{Q}+}} x(q_i^+)a_r(q_i^+) - \sum_{i=1}^{n_{\hat{Q}-}} x(q_i^-)a_r(q_i^-) + \sum_{i=1}^{n_{\hat{P}}} x(s_i)a_r(s_i) + \sum_{i=1}^{n_{\hat{\Omega}+}} x(\omega_i^+)a_r(\omega_i^+) - \sum_{i=1}^{n_{\hat{\Omega}-}} x(\omega_i^-)a_r(\omega_i^-) = 0,$$

for $r = 1, 2, \ldots, n$, and

$$\begin{aligned} &\sum_{i=1}^{n_{\hat{Q}+}} x(q_i^+)a_{n+1}(q_i^+) - \sum_{i=1}^{n_{\hat{Q}-}} x(q_i^-)a_{n+1}(q_i^-) \\ &\qquad + \sum_{i=1}^{n_{\hat{P}}} x(s_i)a_{n+1}(s_i) + \sum_{i=1}^{n_{\hat{\Omega}+}} x(\omega_i^+)a_{n+1}(\omega_i^+) \\ &\qquad\qquad - \sum_{i=1}^{n_{\hat{\Omega}-}} x(\omega_i^-)a_{n+1}(\omega_i^-) \quad = 1 \end{aligned}$$

$$x(q_i^+) \geq 0, \quad \text{for} \quad i = 1, 2, \ldots, n_{\hat{Q}+},$$

$$x(q_i^-) \geq 0, \quad \text{for} \quad i = 1, 2, \ldots, n_{\hat{Q}-},$$

$$x(s_i) \geq 0, \quad \text{for} \quad i = 1, 2, \ldots, n_{\hat{P}},$$

$$x(\omega_i^+) \geq 0, \quad \text{for} \quad i = 1, 2, \ldots, n_{\hat{\Omega}+}, \text{ and,}$$

$$x(\omega_i^-) \geq 0, \quad \text{for} \quad i = 1, 2, \ldots, n_{\hat{\Omega}-}.$$

While the above formulation of the dual is intimidating, we may simplify immediately by recognizing some features of the constraints of our primal problem. We know that the scalar, $b(s_i) \equiv 0$, for all s_i. Then the middle term in the objective function disappears completely.

Let us move to the first constraint. The middle sum also collapses to the single term $x(s_r)$, as we have defined $a_r(s_j)$ to be the Kronecker $\delta(r, j)$. The first three terms of the second constraint disappear altogether, as we have specified, $a_{n+1}(s_j) = a_{n+1}(s_j) = 0$, for all j. The non-negativity constraints remain the same.

E. A SAMPLE SOLUTION

We now use the formulation of the image reconstruction problem as a linear program to solve a simple problem. We first discuss the geometry of the partition that we are using, and then identify some additional simplifying assumptions that make the problem more tractable. We introduce the expected density of our sample problem, and conclude with the Simplex solution of the problem.

1. The Partition

The partition that we use in this example is illustrated in Figure 16, where the color of a polygon is a function of its area. Larger values correspond to

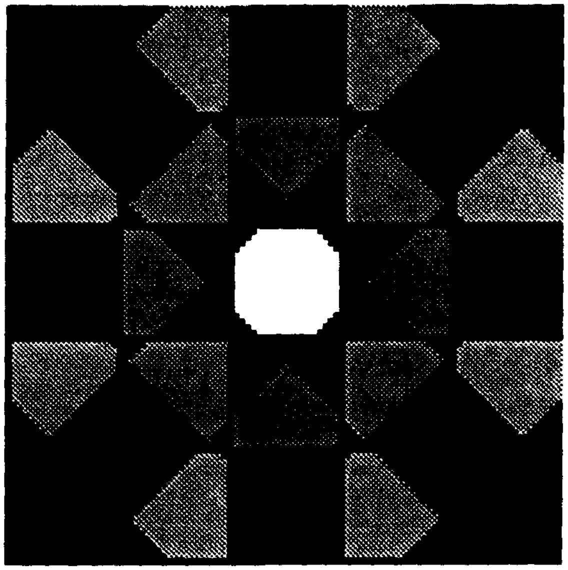

Figure 16. *The Partition of the Sample Problem*

lighter colors. We have chosen the four angles $0, \frac{\pi}{4}, \frac{\pi}{2}$, and $\frac{3\pi}{4}$, with five strips for each angle. The resulting partition consists of 89 polygons.[1] It should be noted that each strip has width of $\frac{1}{5}$. Consequently, as views at angles of $\frac{\pi}{4}$ and $\frac{3\pi}{4}$ require more than 5 strips to cover the unit square completely, only the portion of the square that falls in the five center strips is considered. The rest is ommitted.

2. A Simplifying Assumption

Rather than attempt to solve the infinite dimensional problem as derived in the initial portion of this chapter, we project the target density onto the n-dimensional polygonal basis of our partition. That is, we insist that the target

[1]The manner in which the polygons were identified and the areas of each polygon computed is not a matter of particular concern here. It is sufficient to state that the symmetry of the partition was deeply exploited in a manner which simplifies the problem of polygon identification and area computation over 89 polygons to one of many fewer than 89.

function be constant on each of the polygons of the partition. This simplification reduces the infinite index sets [illegible] and Ω^- to finite sets, as we need only consider a representative $\omega_j \in s_j$ for each polygon s_j when determining the norm of the difference between our target function and optimal function. Without this assumption, the problem is very similar to the infinite problem of approximating the exponential with polynomials, which was discussed in more detail when the Simplex algorithm was introduced. It is possible that this problem [illegible] solvable without this simplification, but no attempt is made to solve it in this thesis.

We choose, in projecting the target function onto our finite dimensional space, the density of the function over each polygon divided by the area of the polygon. That is, after the target function g is projected into the finite space, it takes on the form:

$$\hat{g} = \sum_{j=1}^{n} \beta_j \psi_j,$$

where

$$\beta_j = \left(\frac{\int_{s_j} g dA}{\int_{s_j} dA} \right).$$

That is, $\beta \equiv$ the mass of g over polygon s_j, divided by the area of polygon s_j.

3. The Target Function

We now identify the target function of the sample problem. We use the simple function,

$$g(x,y) = \left(x - \frac{1}{2}\right)^2 + \left(y - \frac{1}{2}\right)^2,$$

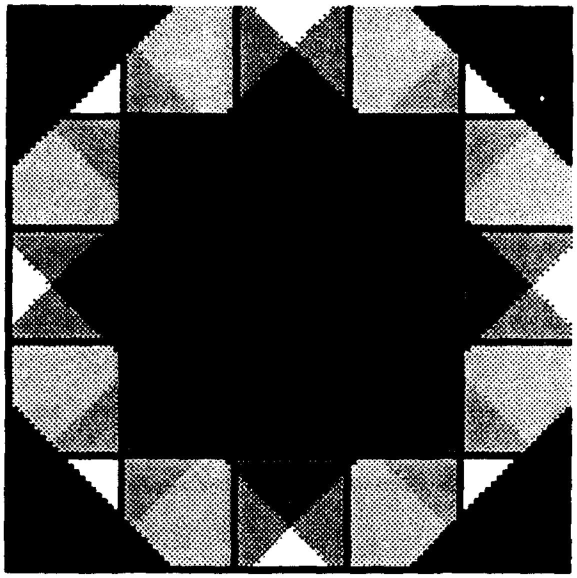

Figure 17. *The Projected Target Function*

for $x \in [0,1]$ and $y \in [0,1]$. The projection of the density function is illustrated in Figure 17. The particular data for the constant densities assigned to each of the strips represent the values which we hope, or expect, to find in the solution of our problem, before considering the data. That is, the values are assumed to represent the most likely solution to our problem.

4. The Test Density

The density function that we use to generate the test data is given by the expression

$$h(x,y) = \begin{cases} \left(x-\frac{1}{2}\right)^2 + \left(y-\frac{1}{2}\right)^2, & \text{for } \left(x-\frac{1}{2}\right)^2 < \frac{1}{16} \\ 1, & \text{otherwise.} \end{cases}$$

The density function is displayed in Figure 18. The values of sample transform defining integrals become the right hand side of our equality constraints when we

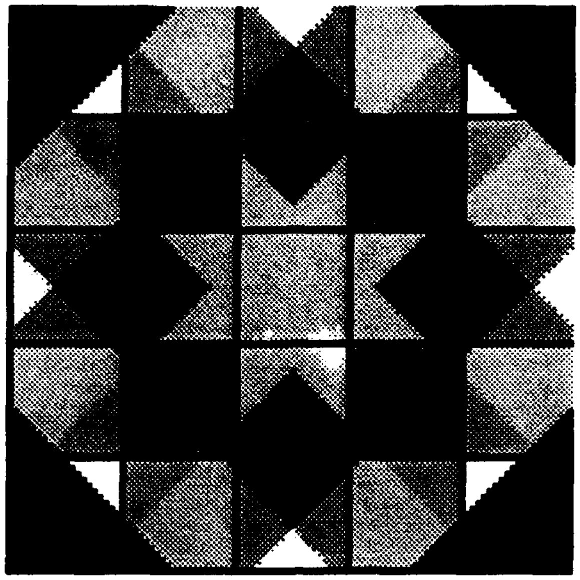

Figure 18. *The Test Density Function*

formulate the problem.

The manner in which we have defined the projection of a density onto the finite dimesional space assures us that both the test density h and its projection have the same sampled transform. Thus, barring catastrophic rounding error, the formulation is always feasible. That is, there must be some density function that produces the sampled transform, even after we have projected the test density onto the partition. If the sampled transform is uniquely determined, we reconstruct the projection perfectly, though the value of the variable k may be quite large.

As a basis of comparison, we note that the maximum difference between the projection of the target density and the test density is given by $d = .1949$. We may certainly expect then, that the optimal density varies by no more than the above value of d.

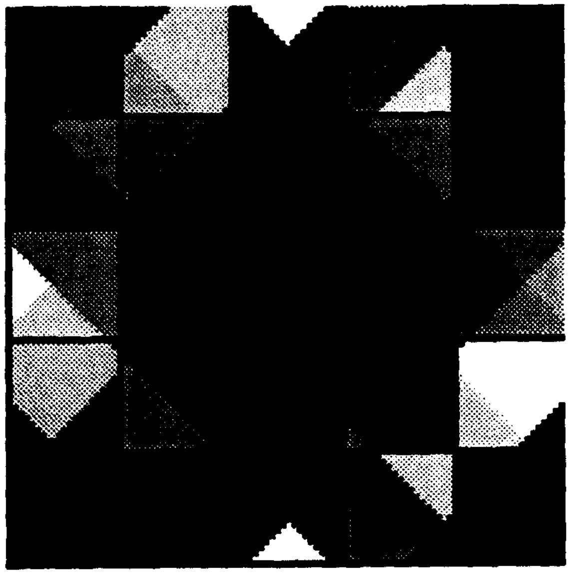

Figure 19. *The Simplex Solution of the Sample Problem*

The optimal density as determined by the Simplex algoritm is displayed in Figure 19. The value achieved for the maximum absolute deviation between the target density and the optimal density is $d' = .1577$. We consider the difference over each polygon in Figure 20.

F. SUITABILITY OF SIMPLEX IN IMAGE RECONSTRUCTION

We briefly consider the merit of using the Simplex algorithm to solve the image reconstruction problem. That is, we wish to consider how well the tool we have chosen fits our particular job.

The results of this particular example show the tendency of this formulation to spread error over the entire region. This consequence, it is believed, results from the use of the infinity norm. We may also question the choice of target functions,

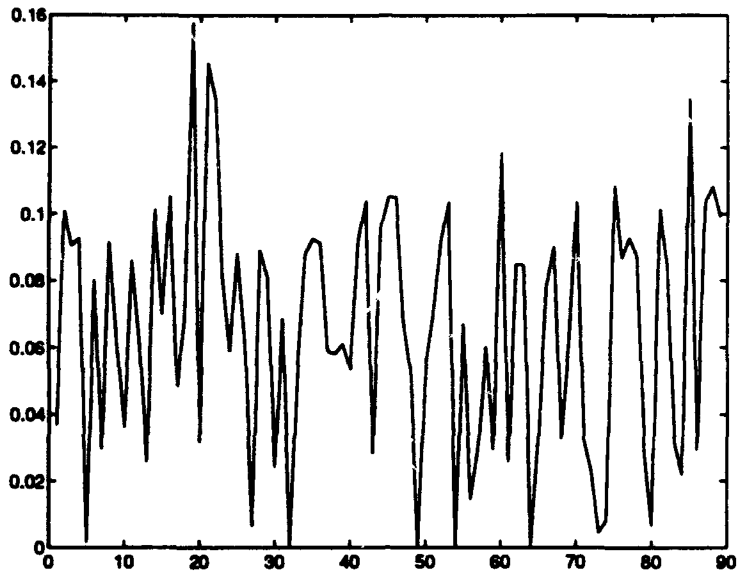

Figure 20. *Difference over Each Polygon*

and may look to other methods of qualification. However, the optimization problem achieves what it is designed to achieve. That is, we have found the density that satisfies the minimum deviation in the uniform norm sense.

We are also forced to consider the substantial data that are required to solve the problem. The problem of polygon identification is a difficult one by itself, especially in view of the fact that a typical partition for the CAT scan problem is generated by 200-300 angles with up to 500 strips per angle. With this geometry, we know that the number of polygons exceeds 4,000,000. Further, we require the area of each polygon be known to solve the problem as we have formulated it. Finally, as each polygon gives rise to a variable in our primal problem, we are solving a problem in a subspace of $\mathcal{R}^n$ where n is quite large. On the positive side, we know that we must solve the polygon identification and polygonal area problems only once. Further, the matrix,

A which results from the formulation above is extremely sparse, which may lead to a more rapid solution of the Simplex problem, or invite other methods of solving Linear Programs.

In conclusion, the author contends that the Simplex algorithm fits well conceptually, but may not suited for the vastness of the problem as it is formulated here. Projecting the density functions onto the polygonal partion is conceptually identical to selecting finite subsets of an infinite index set. The theorems presented in regard to the image reconstruction problem indicate that we may solve the infinite dimensional problem through a sequence of finite dimensional problems, when certain conditions are met.

Some alternatives that might warrant future consideration are norms other than the infinity norm, or using the Simplex model to refine existing solutions to the reimaging problem, where the number of variables is less restrictive.

VII. CONCLUSION

In conclusion, we have introduced the Simplex algorithm in a context quite apart from its usual applications. The principal vehicles for the exploration of the algorithm were three unique applications, each illuminating distinct features of the theory underlying the implementation of the Simplex algorithm.

In particular, we began with a problem of finding orthogonal monic polynomials over a closed interval. This example led to a very basic Simplex formulation, and was solved as a finite dimensional problem. The requirement that the polynomials be monic facilitated the relatively simple formulation. Follow on problems to this example might be the adaptation of the algorithm to generate an non-polynomial orthogonal basis for an infinite dimensional function space, or perhaps to fit the algorithm to solving the non-linear orthonormal basis generation problem.

Second, we formulated the problem of approximating a function over a closed interval in the uniform norm sense. Unlike the first example, the problem was infinite dimensional, in that the formuation required a constraint for every number in the uncountably infinite set. This problem proved particularly helpful in illustrating the principle of weak duality, and ultimately illustrated the Simplex algorithm itself. The special qualities of polynomial approximation were ommitted, though the reader is referred to [Ref. 9] for a more complete discussion thereof. Again, potential areas for future research might include approximation with functions other than polynomials.

Each of the above examples were used extensively to illustrate the highlights of convexity and duality, upon which the Simplex algorithm is based. The treatment was relatively general, though many of the theorems required that the linear optimization problem be finite. Work is underway to identify classes of infinite dimensional problems which may be solved by a sequence of finite dimensional problems. The reader is referred to [Ref. 11] for more complete discussion of this active area of research. Highlights include infinite horizon planning, fuzzy set semi-infinite programming, and linear programming in control theory.

Another area of focus in this paper was on the Image Reconstruction problem. Again, this is an area of active research. After presenting the requisite background, we formulated this problem as an infinite dimensional linear optimization problem, and as a special case of the uniform approximation problem. Results were presented that indicated that use of Simplex to solve a sequence of linear programs is conceptually sound, though not necessarily practical in view of the size of the problem. This application of the algorithm, however, is open to more extensive research in a number of areas. A different choice of norms by which the quality of density functions is measured may eliminate a number of constraints. A technique for formulating optimization problems with the 2-norm is found in [Ref. 12], and may prove useful in this application. The Simplex algorithm may also provide an inexpensive method to solve coarser problems, from which one may determine the necessity of constructing more detailed models. Alternately, there may be some utility in using the algorithm to solve the reconstruction problem in only small portions of the set over which a density is defined. If there is utility in such an application, the logical consequence is research of parallel Simplex implementation.

The potential utility of the Simplex algorithm to unconventional applications seems clear. Even when actual implementation of the algorithm is not practical, the tools of convexity and duality apply to broader areas of optimization.

REFERENCES

[1] James P. Keener. *Principles of Applied Mathematics.* Addison-Wesley Publishing Company, Reading, Mass., 1988.

[2] Jack Indritz. *Methods in Analysis.* The Macmillan Company, New York, N.Y., 1963.

[3] Stanley R. Deans. *The Radon Transform and Some of its Applications.* Wiley Interscience, New York, N.Y., 1983.

[4] Bruce T. Robinson. *A Multilevel Approach to the Algebraic Image Reconstruction Problem.* PhD thesis, Naval Postgraduate School, 1994.

[5] Mark A. Limber and Stephen F. McCormick. *Optimal Resolution in Maximum Entropty Image Reconstruction form Projections with Multigrid Acceleration.* Unpublished Note, 1994.

[6] G.T. Herman and S.A. Johnson. Image reconstruction from projections. *Scientific American*, 233:56–68, 1975.

[7] Mokhtar S. Bazaraa and Hanif D Sherali. *Linear Programming and Network Flows.* John Wiley and Sons, New York, N.Y., 1990.

[8] Manfred Walk. *Theory of Duality in Mathematical Programming.* Springer-Verlag Wien, New York, N.Y., 1989.

[9] Klaus Glashoff and Sven-Ake Gustafson. *Linear Optimization and Approximation.* Springer-Verlag, New York, N.Y., 1983.

[10] Josef Stoer and Christoph Witzgall. *Convexity and Optimization, I.* Springer-Verlag, New York, N.Y., 1970.

[11] A.V. Fiacco and K.O. Kortanek, editors. *(Proceedings of the Symposium on) Semi-Infinite Programming and Applications.* Springer-Verlag, New York, N.Y., 1981.

[12] Ralph E. Steuer. *Multiple Criteria Optimization Theory, Computation, and Application.* Wiley Interscience, New York, N.Y., 1986.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center — 2
   Cameron Station
   Alexandria, VA 22304-6145

2. Library, Code 52 — 2
   Naval Postgraduate School
   Monterey, CA 93943-5002

3. Department Chairman, Code MA — 1
   Department of Mathematics
   Naval Postgraduate School
   Monterey, CA 93943-5002

4. Prof Van Emden Henson, Code MA/Hv — 1
   Naval Postgraduate School
   Monterey, CA 93943-5002

5. Prof Robert F. Dell, Code OR/De — 1
   Naval Postgraduate School
   Monterey, CA 93943-5002

6. LT Christopher A. Rhoden — 4
   1267 Nutwood Drive
   Centerville, OH 45458